A first principles study of proton transport through model helical pores

Lynsey M. S. Shepherd

Doctor of Philosophy

University of Edinburgh

June 2013

Preface and Declaration

Since graduating from the University of Edinburgh in 2007 with an MChem (Hons) degree in Chemistry with a year in Industry the author has been engaged in a program of full time research under the supervision of Dr Carole. A. Morrison at the University o Edinburgh.

No part of the work referred to in this thesis has been submitted previously in whole or in part for another degree or qualification from this or any other university or institute of learning. In accordance with the regulations this thesis does not exceed 70,000 words in length.

19 05. 2013

Preface and Declaration

Since graduating from the University of Edinburgh in 2007 with an MChem (Hons) degree in Chemistry with a year in Industry the author has been engaged in a program of full time research under the supervision of Dr Carole. A. Morrison at the University of Edinburgh.

No part of the work referred to in this thesis has been submitted previously in whole or in part for another degree or qualification from this or any other university or institute of learning. In accordance with the regulations this thesis does not exceed 70,000 words in length.

Abstract

Proton transport (PT) across cell membranes is a fundamental process and a key step in many biological functions, including cell signalling and enzymatic reactions. All biochemical reactions that convert energy from one form to another are mediated by PT, which also serves as a vital route to achieve cell pH stabilisation. The coding for membrane-bound proteins constitutes 25-30% of all genes, and they are implicated in many diseases such as diabetes and Parkinson's. Consequently, they are the subject of major drug target studies (in fact the drug targets for all neurological diseases are membrane-bound proteins). Whilst PT is known to occur via transient water molecules across the cell membrane itself, it is more often the case that the mechanism involves proteins that span the membrane surface and act as proton-specific ion channels. PT has been widely studied in protein systems such as gramicidin A, cytochrome C oxidase, the M2 channel protein in the influenza A virus and bacteriorhodopsin. Evidence for the relay of H^+ by buried water molecules ('water wires') mediated by the side-chains of alpha-helices have been substantiated in these and other proteins, but finding direct experimental evidence for the reaction pathway is extremely challenging work.

When experiment can provide only partial answers, it is the role of computational modelling to complete the picture. Modelling these trans-membrane proteins at the full atomistic quantum mechanical level, however, lies beyond the capabilities of current computational techniques, necessitating the use of simplified models. To this end, work undertaken in this thesis has derived and tested a simplified model that is large enough to maintain the essential tertiary structures of transmembrane proteins, but small enough to permit full *ab initio* MD simulations over long time periods to be performed. The model is based on a single helix scaffold placed under periodic boundary conditions to create a cavity that supports a water wire. The simulations then focus on monitoring the behaviour of a proton as it 'hops' along this wire in a manner akin to the classical Grotthuss mechanism.

Mechanistic studies have taken place using poly-glycine, poly-glycine-serine and poly-glycine-aspartic models, and show that the mechanism of PT in channel environments shares some features with the simulations reported for bulk water, with,

e.g., the hydrogen bond distance shortening in the time period leading up to successful proton transfer. There are, however, also some important differences, such as the observation of a heightened number of proton rattling events. The channel environment also removes the need for the loss of a water molecule from the inner coordination sphere of the receiving water molecule as the constriction in space only allows a coordination sphere of three molecules, as opposed to four for bulk water.

The effect of varying the density of water molecules in the channel has also been investigated. A range of cationic states have been identified, with widely varying lifetimes and compared across all models. We also observe that the helix plays an important role in directing the behaviour of the water wire: the most active proton transport regions of the water-wire are found in areas where the helix is most tightly coiled. Finally, we report on the effects of different DFT functionals to model a water-wire using the simplest poly-glycine model, and on the importance of including dispersion corrections to stabilize the helical structure.

Finally, using the poly-glycine-aspartic acid model, a study was undertaken that focused on the direction of proton transport through the channel when the side chains of the aspartic acid residues interacted directly with the water wire. In this model there were two different pathways for the excess proton to pass along: a long hydrogen-bonded network of water molecules and amino acid residues, or a short $[H_3O]^+$ diffusion pathway. It was found that the proton-hopping route over multiple water molecules and amino acid residues was preferred over the diffusion route, even though this pathway was substantially longer.

Acknowledgements

First and foremost I would like to thank my supervisor, Dr. Carole A. Morrison for her enduring patience! Without her constant guidance, immense knowledge and unwavering (almost) belief that I would finish, well, I definitely wouldn't have.

Second I have to thank Morrison group members past – specifically Martin Walker, David 'Manuel' Martins and Anthony Reilly for their patience with my never-ending questions....I never said I could program. The rest of the office crew also get a massive thanks - especially Matt, Dean and Michael for their hair-brained schemes (and yes again help with programming!): you guys kept me entertained always.

To all those involved in cake Thursdays nay Fridays – I thank you for your sweet treats. My waistline does not!

My eternal thanks go to Ben, Zoe and Charlotte for all of the dancing and for keeping me sane.....well to quote Ben – make ME seem sane in comparison! That's the way it is Roach.

Special thanks to Matt, Nic and the Muldoons – you made the last two years in Edinburgh great. And to Dan – for telling me I have to finish! But you are also partly to blame for it taking this long – would I have discovered wine without you? :)

I want to thank my family, particularly Mum and Paul for all of their support over the seven, yes seven, years of university. I could not have done it without you. And Paul, I've finally got a real job!

Table of Contents

Abbreviations

%	percent
Ψ	wavefunction
Σ	sum of
Φ	basis functional
α	spread of a Gaussian function
δ	change in
∇	Laplacian Operator
ε_0	vacuum permittivity
ħ	Planck's constant divided by 2Π
°	degree
≤	less than or equal to
Å	Angström
a	acceleration
a, b, c	cell vector
ab initio	from the beginning
ARF	Atomic Radial Distribution Function
ASP	Aspartic Acid
Asn	Asparagine
B3LYP	Becke's three parameter exchange functional and Lee, Yang and Parr's non-local correlation functional
BLYP	Becke's exchange functional and the non-local correlation functional of Lee, Yang and Parr.
BFGS	Broyden-Fletcher-Goldfarb-Shanno method of geometry optimisation
BO	Born-Oppenheimer approximation
BR	Bacteriorhodopsin

c	weighting co-efficient
ca.	*circa* (approximately)
CcO	Cytochrome C Oxidase
CPMD	Car-Parrinello Molecular Dynamics
CP2K	Molecular Dynamics program used throughout this thesis
DFT	Density Functional Theory
e	charge of an electron
e.g.	for example
et al.	*et alli* (and others)
etc.	*etcetera* (and so forth)
E_{cut}	energy cut off for plane wave basis sets
eV	electron volts
F	force
fs	femtoseconds
gA	Gramicidin A
GGA	Generalised Gradient Approximation
Glu	glutamic acid
GLY	glycine
$\hat{H}$	Hamiltonian operator
H-Bond/ed/ing	hydrogen bond/ed/ing
HF	Hartree-Fock
i.e.	*id est* (that is)
k	wavevector of the de Broglie wave OR force constant
$kJmol^{-1}$	kilojoule per mole
LCAO	Linear Combination of Atomic Orbitals
LDA	Local Density Approximation
m	mass

MD	Molecular Dynamics
MM	Molecular Mechanics
MP2	truncated second order Møller-Plesset perturbation theory
MS-EVB Theory	Multi State – Empirical Valence Bond Theory
n	number
NVE	Microcanonical ensemble with a constant number of atoms, volume and energy
NVT	Canonical ensemble with a constant number of atoms, volume and temperature
O*	defined as the oxygen which carries the excess proton, oxonium ion, $[H_3O]^+$
O_{next}	Defined as the oxygen which will accept a proton as a result of proton transport
$O_{nearest}$	defined as the other nearest neighbours to the oxonium ion
PBC	Periodic Boundary Conditions
PBE	The DFT functional of Perdew, Burke and Ernzerhof
PES	Potential Energy Surface
poly-GLY	poly-glycine model
poly-GLY-ASP	poly-glycine-aspartic model
poly-GLY-SER	poly-glycine-serine model
ps	picoseconds
PT	proton transport
PW91	The DFT functional of Perdew and Wang formulated in 1991
QM	Quantum Mechanics
QM/MM	Quantum Mechanics / Molecular Mechanics
r	position
RMSD	Root Mean Square Deviation
t	time
TD-RDF	Time Dependent – Radial Distribution Function

TMP	Trans-membrane Proteins
υ	velocity
via	by way of
VMD	Visual Molecular Dynamics Package
$\dot{x}$	reaction co-ordinate of interest *e.g.* bond length
XRC	X-Ray Cystallography
Z	atomic number of an atom
ZPC	Zero Point energy Contributions

Chapter 1:

The Poly-glycine Model

An introduction to proton transport theory and the simulation techniques used to model it

1.1 Proton Transport in water

Proton transport (PT) across cell membranes is of fundamental importance in chemical biology, giving rise to many basic functions including bioenergetics, cell signalling and pH regulation.[1-3] Whilst PT is known to occur *via* transient water molecules across the cell membrane itself, it is more often the case that the mechanism involves proteins that span the membrane surface and act as proton-specific ion channels. Evidence for the relay of H^+ by buried water molecules ('water wires') mediated by the side-chains of alpha-helices has been substantiated in several channel and signalling trans-membrane proteins (TMPs). Specific systems include the ion channel protein bacteriorhodopsin [4], the influenza A virus M2 protein [5], the MotA-MotB protein component of the flagella that drive E. Coli [6], and the aquaporins.[7] The coding for membrane-bound proteins constitutes 25-30% of all genes, and they are implicated in many diseases such as diabetes and Parkinson's. Consequently, they are the subject of major drug target studies (in fact the drug targets for all neurological diseases are membrane-bound proteins). Their importance was further recognised in 2003 when the Nobel Prize for Chemistry was awarded to Agre and MacKinnon for their discoveries concerning channels in cell membranes.

Finding direct experimental evidence for PT is extremely challenging work. The protons are thought to be carried by a chain of water molecules, but this level of detail lies beyond the current limit for experiment. Electrophysiology experiments can provide a qualitative picture of the conduction properties of a channel by voltage measurement, but it is difficult to obtain a clear picture of what is happening. Crystal structures can also provide insight, but it too is limited. To date the most complete atomistic models available for a TMP are those for bacteriorhodopsin and its analogue, bovine rhodopsin, both of which are applied *via* homology modelling as general models for membrane-bound drug target proteins.[8-10] The bacteriorhodopsin structure was obtained from high-resolution electron microscopy data supplemented with X-ray diffraction data of 1.55 Å resolution.[11] Whilst this is the current state-of-the-art it is still possible to miss-assign amino acid residues, and hydrogen atom positions are far beyond reach. In addition concerns exist as to whether the protein is captured in its active or passive form. All of these issues have critical implications in understanding how the protein functions from an atomistic level.

When experimental methods yield only partial results it is the role of theory to complete the story. The accurate modelling of PT events also presents its own challenges as the mechanism involves bond making/breaking events, which requires that a high level of computational modelling such as quantum mechanics (QM) is employed. In addition hydrogen atoms may quantum tunnel, and along with zero-point energy effects this may radically alter the reaction landscape. To further compound these problems, the proton hop step is a borderline rare event, meaning that some computational effort will be 'wasted' on the uninteresting wait time between events, resulting in obvious problems with statistical sampling. As with all simulations in chemical biology, a balance must be struck between a realistic (*i.e.* atom-rich) model that can only be supported by a low-to-medium level of computational accuracy [such as classical molecular modelling (MM) and its extensions MS-EVB [12-14] and QM/MM [15, 16]] and a simplified, less realistic, atom-light model, that can be subjected to the higher-level quantum mechanical-based methods. Examples of the first all-atom approach are plentiful in the literature, with most attention understandably focused on small, structurally well defined TMPs such as bacteriorhodopsin[17-19], the M2 channel of the Influenza A virus [20, 21] and cytochrome C oxidase.[22-25] Another example is the use of the synthetic leucine-serine channel (LS2) [26] to study PT, although this model is still too large to allow for full QM methods.[27, 28] We stress that if the PT mechanism is suspected to involve a water-wire system then it is important that the simulation takes this into account, rather than opting for the simpler unprotonated water chain system as their behaviours can be markedly different.[29]

The simplest system that can support proton transport (PT) is a water dimer; protons are exchanged between donor and acceptor atoms that form a hydrogen bond. When the distance is greater than *ca.* 2.5 Å the potential energy surface (PES) comprises a double well [see Figure 1.1(a)]. Under classical conditions PT is achieved when the proton 'hops' over the transition state. The greater the inter-oxygen distance the higher the barrier to PT becomes, meaning that, at longer bond lengths it becomes increasingly more difficult for the proton to obtain the energy to 'hop' to the other side.

Conversely, as the O...O interaction shortens (and the hydrogen bond becomes stronger) the activation barrier is reduced until, at distances of < 2.4 Å, it may disappear altogether to form a 'bath-tub' like PES [see Figure 1.1 (b)].

This facilitates the movement of the proton from one side to the other and illustrates the dependence of the PT mechanism on donor-acceptor separation. As this distance is modulated by the vibrational motion of the molecules this suggests that PT is reliant on this motion, with a coupling between O-H and O...O stretching modes required to gain the correct configuration for transfer to occur.

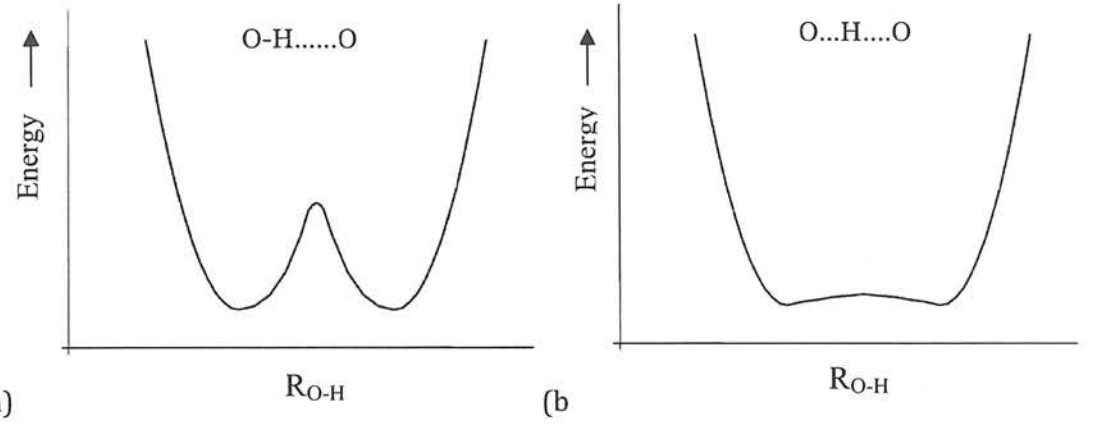


Figure 1.1 (a) The hydrogen bond potential energy surface when donor-acceptor separation >2.4 Å and (b) donor-acceptor separation >2.4 Å

Hydrogen has the smallest and lightest nucleus and as such is subject to quantum effects, such as tunnelling. Tunnelling is the phenomenon whereby a particle is able to move through a classically forbidden energy state – instead of having to 'hop' over the activation barrier, the proton can simply cross 'through' it. In the context of PT this means the shortening of the O-O distance is not always necessary, and PT can occur over distances greater than 2.4 Å.

Of course the case of proton transfer in bulk water is more complicated than in a simple dimer. In bulk water an excess proton exists as a larger solvated structure. Two types of solvated excess protons are widely reported in the literature, namely Zundel and Eigen complexes. The Zundel complex (or dihydronium ion) consists of two water molecules and an additional proton to form the $[H_5O_2]^+$ cation [Figure 1.2(a)]. It is characterised by a short strong hydrogen bond formed between the two oxygens to give a O-O distance of ~ 2.4 Å. This means that the bonding is symmetric in nature, and rather than an explicit acceptor-donor scheme the proton sits almost exactly in the middle of the two oxygen centres. The Eigen cation is larger, with three additional water molecules forming hydrogen bonds around a central $[H_3O]^+$ [see Figure 1.2(b)].

Three distinct hydrogen bonds are formed between each hydrogen of the oxonium and acceptor oxygens on the water molecules to give the $[H_3O.3H_2O]^+$ or $[H_9O_4]^+$ cation. Simulations of an excess proton in bulk water have suggested an Eigen to Zundel ratio of 65:35, with the Eigen being the more stable of the two.[12] The interplay between these two structures is thought to be integral to the mechanism of fast proton transport in this medium.

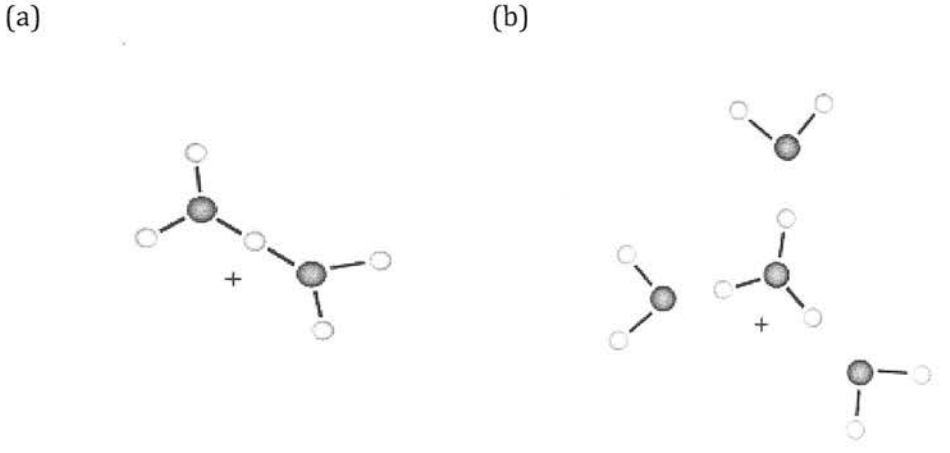


Figure 1.2 (a) Zundel cation, $[H_5O_2]^+$ and (b) Eigen cation, $[H_9O_4]^+$

Several different mechanisms have been postulated for the behaviour of excess protons in bulk water. The oldest and most widely accepted is the 'Grotthuss Mechanism,', originally postulated over 200 years ago.[30] Here the proton can simply 'jump' along a hydrogen-bonded chain of water molecules. An important feature of this mechanism is that the identity of the proton changes as it moves along the chain, showing that it is not just simple diffusion of H_3O^+. A schematic diagram of the Grotthuss mechanism is shown in Figure 1.3; in essence a proton enters from the left and hydrogen bonds to a water molecule to form an H_3O^+ cation. A different proton from this hydronium is then transported to a subsequent water molecule in the chain *via* an existing hydrogen bond. This continues along the length of the chain. The critical feature of this mechanism is that as the identity of the transferring proton changes, an excess proton is able to diffuse throughout the entire hydrogen bonded network of water at a rate considerably higher than that of conventional diffusion of the H_3O^+ species.

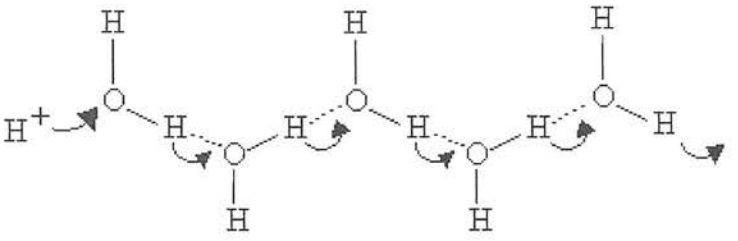


Figure 1.3 Schematic of the Grotthuss mechanism

Of course this is a rather simplistic model, and consequently much work has been undertaken to gain a deeper understanding of the process. It is now known that the solvation state of the oxonium cation in excess water is different from that of a typical water molecule. In bulk water the molecules have a coordination of about 3.9, whereas when an excess proton is present the coordination around the [H_3O^+] is closer to three.[31] This suggests that in order for a PT event to occur there must be considerable rearrangement of the molecules surrounding it.

Agmon [30] proposed a mechanism which requires an Eigen-Zundel-Eigen isomerisation (Figure 1.4). The first step in the process involves the shortening of the O...O bonded distance to form a symmetrical hydrogen bond motif (Zundel complex). Before transfer can occur the accepting water molecule (termed O_{next}) must lose a hydrogen bond, dropping its coordination number from four to three as an Eigen cation is formed. The water molecule left behind after transport gains a hydrogen bond, so its coordination number increases from three to four. This has been called the 'Moses mechanism' because of the similarity with Moses parting the Red Sea.

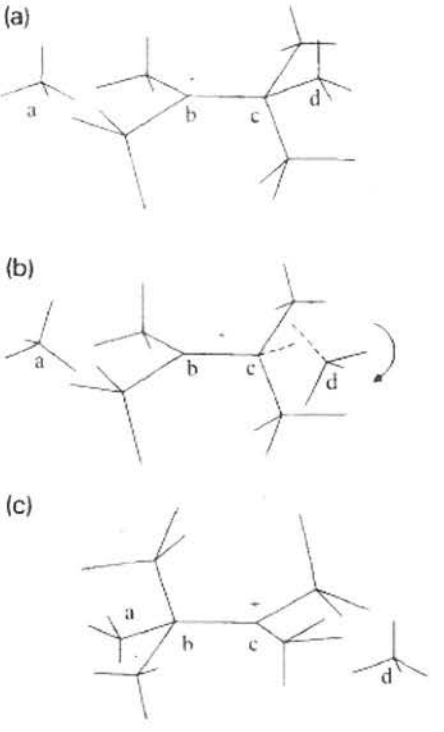


Figure 1.4.The Moses mechanism for PT in aqueous environments. The proton is localised on oxygen b [Figure 1.4(a)] A hydrogen bond between Oc and Od is broken, facilitating the formation of a Zundel complex between Ob and Oc [Figure 1.4(b)]. Finally, in figure 1.4(c), the proton is transferred from Ob to Oc and a new hydrogen bond is formed between Oa and Ob.

Lapid *et al* [32] have offered a more complicated version that involves the breaking of multiple bonds within the extended hydrogen-bonded network, rather than the single interaction, as considered in the Moses mechanism. The multiple bond breakages serve to disrupt the symmetry around the excess proton and stimulate PT, thereby acting as a collective process (see Figure 1.5). Many studies have tried to prove which of the two mechanisms is correct – particularly the groups of M. Parinnello [33-35] and G. A. Voth. [3, 36, 37] The validity of either method has yet to be proven, however recently it has been suggested that the true mechanism is likely to be the more complicated extended version proposed by Lapid.[3] Recent simulations on bulk water by Berkelbach [38] have used Time-dependent Radial Distribution Functions (TD-RDFs) of the oxonium oxygen (O*), the receiving oxygen (O_{next}) and the two nearest neighbours ($O_{nearest}$) to probe the true mechanism. These show that there is indeed a

change in the coordination number of the O* from three to four and conversely the O_{next} from four to three after transfer. The RDFs also show a shift to shorter O...O separation as the point of transfer is approached which indicates the formation of a Zundel cation, $[H_5O_2]^+$, although they do not investigate hydrogen bond behaviour outside of the first solvation shell.

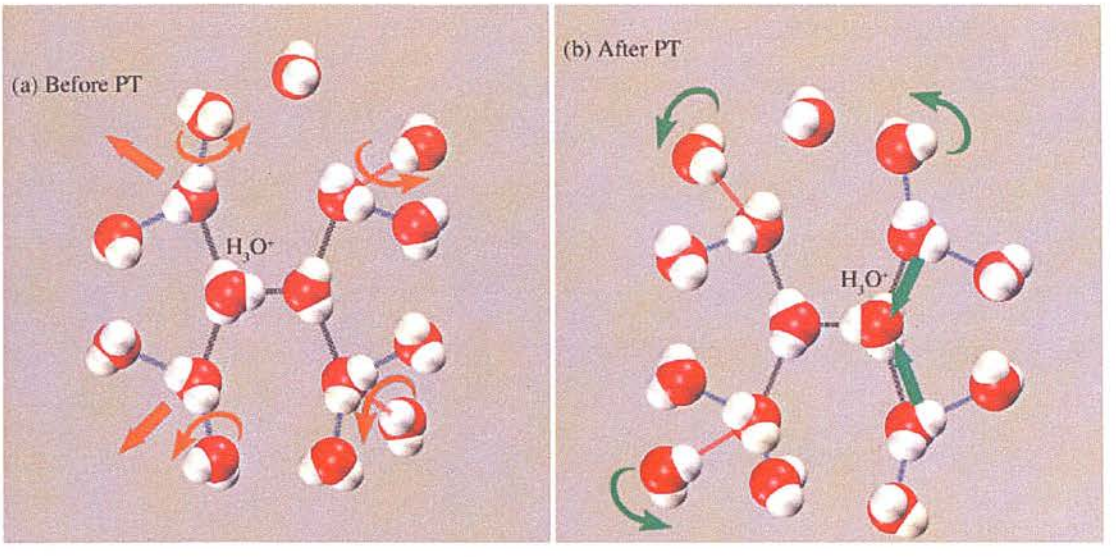


Figure 1.5 Lapid's mechanism for PT in an aqueous environment. (a) Before PT. The orange arrows represent hydrogen bonds breaking in the first and second solvation shells. A Zundel complex is formed which then breaks down to form the new $[H_3O]^+$ (b). After PT. Hydrogen bonds reform (green arrows) in the first and second solvation shells.

1.2 Proton transport through channel environments

The next step in the path to simulating PT in a protein channel is to consider a restricted environment that encapsulates chains of hydrogen-bonded water molecules. Numerous models have been used that encompass smooth-walled hydrophobic channels, synthetic proton channels and explicit models of ion channels.

1.2.1 Smooth-walled channels

Simplified models have in the past provided valuable insight into the process of PT through membrane-bound proteins, and offer the considerable benefit that their simplicity allows generic principles to be obtained. Restraining potentials *in vacuo* have been used to create hypothetical cylindrical pores into which a chain of water molecules are confined. Pomès and Roux [39, 40] used constraints on the water oxygen atoms to mimic a hydrophobic pore which supports wires of different lengths. The resulting PT steps were consistent with the general idea behind the Grotthuss mechanism, *i.e.* the proton 'hops' along a hydrogen-bonded chain of water molecules and the identity of the proton changes with each transfer step. Brewer *et al* [41] also created a non-polar pore using a repulsive potential. Both Eigen and Zundel cations were observed, with the Eigen cation more prevalent in the wider channels of the study where there was room to accommodate it. The formation of the Eigen complex, $[H_9O_4]^+$, means that the direction of PT is no longer limited to the channel axis; it can be viewed as an 'ion trap'. As the channel width decreases the stabilities of the Eigen and Zundel cations become almost identical so that inter-conversion between the two is easier and the rate of PT increases. At very narrow pore radius a continuous proton wire is formed and, as there is no potential for Eigen formation, there is no possibility of a proton 'trap'.

Hummer [42] has looked at proton transport in the context of carbon nanotubes as a model for a non-polar pore (see Figure 1.6). A carbon nanotube and internal central water-wire were simulated using both Car-Parrinello molecular dynamics (CPMD) and Multi-State Empirical Valence Bond theory (MS-EVB) and its derivatives. The work allowed for proton diffusion constants to be calculated; the rate of transport through the nanotube *via* the water-wire was an order of magnitude higher than those calculated for transport in bulk water using MS-EVB.

Figure 1.6 A solvated carbon nanotube with the internal water molecules forming the water wire shown in CPK space filling notation.

Hummer has also sought to investigate the rate-limiting step in PT using a model where a proton-wire has not been constructed.[43] Instead a chain of water molecules has been placed within the carbon nanotube and the excess proton outside of the nanotube in the bulk. In order for a wire to be formed, the excess proton must diffuse into the nanotube from the reservoir. The entrance to the channel does appear to be integral to the rate of transport; in this case the proton did not enter the channel during the simulation time. The free energy profile of proton position along the length of the nanotube (z-axis) was also calculated. The movement along the wire itself is almost barrier-less and indicated that, in this case, the probable barrier to transport is the desolvation penalty of moving from the bulk to the wire.

A further paper showed water quickly diffusing into empty nanotubes from the water reservoir.[44] This is somewhat counterintuitive, as the potential for stabilising hydrogen bond formation is reduced in the pore. In bulk water each water molecule forms four hydrogen bonds, whereas in the pore environment a maximum of two is possible. There is some stabilisation provided in the pore due to carbon interactions (modelled *via* a Lennard-Jones potential). If these interactions are removed the channel

will remain empty, with only transient wires formed. This paper also offers an interesting insight into the rate of PT in a channel environment *versus* bulk water; in the channel PT is a localised process and does not require co-operative fluctuations of hydrogen-bonds as far away as the second solvation shell.

1.2.2 Synthetic ion channels

A synthetic leucine-serine channel has been used as a more realistic model of the interior of a transmembrane protein, *i.e.* a channel which is hydrophilic and whose size and shape varies along its length. As in the hydrophobic channels Eigen complexes are formed in the wider channel regions, enforcing the theory that they can form a barrier to PT. The amino acid side chains exhibit strong hydrogen bonding to the water-wire, which suggests their possible involvement in the PT process. [26, 28]

Nagle and Horowitz first suggested the use of hydrogen-bonded chains as a mechanism for proton transport through channels.[45] The amino acid side chains of the protein are folded in such a way as to form a continuous chain, along which protons are able to rapidly diffuse (see Figure 1.7). This has since been overtaken by the more accepted picture of a proton-wire, whereby water molecules are used to bridge 'gaps' between interior polar side chains.

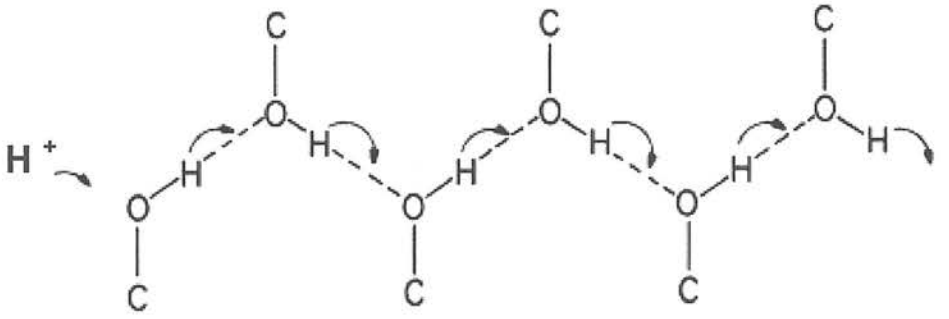


Figure 1.7 Hydrogen bonded chain mechanism of proton conduction

1.3 Transmembrane protein models

With recent advances in computational methodology it is now possible to study the mechanism and dynamics of PT using more complex techniques. Previously most modelling of these systems has been carried out using classical force-field methods such as AMBER [46] or CHARMM [47] to model protein interactions in models containing upwards of 20,000 atoms. Covalent bonds are made and broken during the transfer process and as simple molecular mechanics methods do not model electronic interactions they cannot account for this. Alternative methods are needed and have been developed. QM/MM models a small subsection of a larger model system using quantum mechanical methods while the regions outside the area of interest are modelled classically. Classical force fields have been extended to include special terms to describe the rapidly fluctuating hydrogen bonded structure around an excess proton. This is known as Multi-State Empirical Valence Bond Theory (MS-EVB1,2 and 3).[12, 13, 48] These methods have made it possible to model the pathways of protons through proteins such as gramicidin A, the influenza A virus M2 channel, bacteriorhodopsin and cytochrome C oxidase.

1.3.1 Gramicidin A

Gramicidin A (gA) is an antibiotic found in soil bacteria which forms a very small protein channel by the head to head contact of two fifteen residue polypeptide chains (see Figure 1.8). The interior is lined by inward facing carbonyl backbone groups and as such is hydrophilic. gA is monovalent selective and allows for the transport of cations, such as K^+ and Na^+. The rate of PT through gA is anomalously high, and this, coupled with the absence of significant water flux, suggests a mechanism other than simple hydrodynamic diffusion of H_3O^+ ions. All studies carried out to probe the mechanism of PT through gA find agreement with the modified Grotthuss 'hop' mechanism.[49-52] Voth (using MS-EVB) showed that a continuous water-wire is not necessary as carbonyl groups lining the pore can stabilise the breaks and also help with

the solvation of the excess proton.[52] Pomes and Roux (using QM/MM) observed that PT occurs *via* a 'semi-collective process involving rapid fluctuations of hydrogen bond length in the chain. Here the mechanism could not be exactly defined, but they concluded it was neither concerted nor completely incoherent. [53, 54]

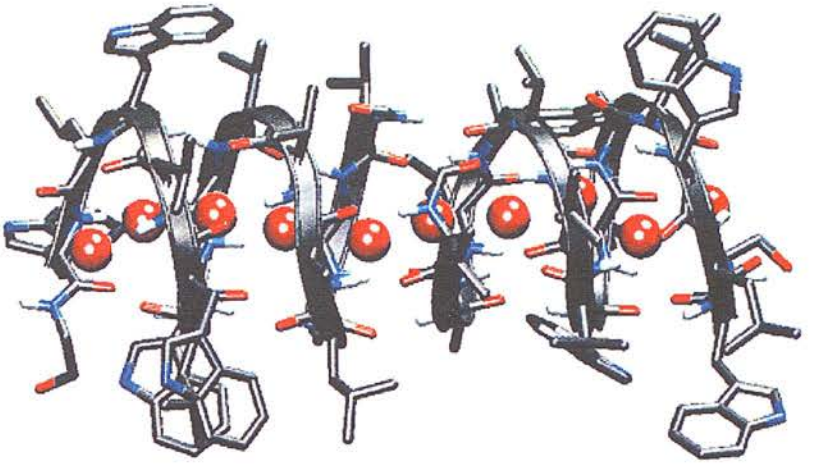


Figure 1.8 Gramicidin A channel with internal water wire.

1.3.2 Influenza A M2 channel

The M2 channel of the influenza A virus is a proton selective channel which is used for the transport of protons for pH regulation. The transmembrane domain of the protein consists of a tetrameric bundle of four nineteen residue alpha helical polypeptide chains (see Figure 1.9). The function of this channel is important for efficient action of the virus; proton flux allows for the uncoating of viral nucleic acid in endosomes which is essential for the virus to enter host cells. [55]

The mode of action for this channel is much debated, but it is thought that a ring of four histidine residues act as a gate to block transfer. Molecular Dynamics (MD) MS-EVB2 simulations on seven different starting configurations were performed by Smondyrev *et al.* [20] Three of the seven showed proton mobility through the channel

(in others the proton was immobilised or diffused back into the bulk). From these simulations three rates of diffusion were calculated and were comparable to those calculated for bulk water. The simulation showed that once again the mechanism of crossing the gated region was not that of simple H_3O^+ diffusion. Protonation of the histidine residues opens the gate and a Zundel complex is formed as part of a transient water-wire to serve as a bridge between the extra- and intra-cellular regions.[56] The importance of the amino acid selection for the gating region has been highlighted using mutagenic experiments. Replacing the histidine residues with alanine, glycine or glutamic acid results in an increase in proton conductance but also a loss in proton selectivity. Substitution with cystine completely stops channel function.[57] More recent work casts doubt over the formation of the transient water wire as the method of proton conduction, however, and would suggest that the protons are moved through by a conformational change of the histidine residues themselves.[21] The results agree with experiment that the triply protonated state is the most likely to be the open form of the channel. MD simulations of this state suggest that although there is space for a water molecule to pass through the gate, the formation of a water wire is unlikely. Transfer is more likely to occur by the release of a proton from one protonated histidine residue. At this time, however, neither mechanism has been ruled out.

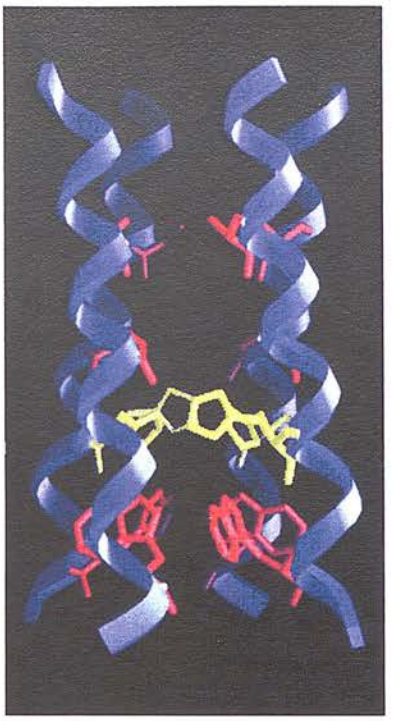


Figure 1.9 The influenza A M2 channel. The histidine residues which form a physical barrier to PT are shown in yellow, with double ringed tryptophan residues shown in pink.

1.3.3 Bacteriorhodopsin

Bacteriorhodopsin (BR) is a small transmembrane protein which serves as a good structural model for the class of ion channels characterised by seven interlocking alpha helices. The small size (27kDa) means that the structure has been well characterised by X-ray crystallography (XRC), with perhaps the best determined by Luecke *et al* [11] at a structural resolution of 1.55 Å, where 223 of the 248 amino acid residues were assigned. XRC, in this application, still has its limitations. Not all residues can be unambiguously assigned and although the presence of internal water molecules is known their positions cannot be exactly found.

The steps in the BR photocycle are well known and accepted and a number of intermediate structures are known (labelled J – O). The pathway is indicted in Figure 1.10, where some key residues named. During step one the Schiff base is deprotonated

and the amino acid residue Asp85 receives a proton. A proton is then released into the extracellular medium (step 2). Asp96 loses its proton which makes its way to reprotonate the Schiff base (step 3), and then a proton is transported from the cytoplasmic side to reprotonate Asp96 which is accompanied by the re-isomerisation of the retinal and a final deprotonation of Asp85. The internal water molecules are thought be involved in the movement of the proton from the Asp96 residue to the Schiff base, and also in the transportation of an excess proton to reprotonate Asp96 from the cytoplasm.

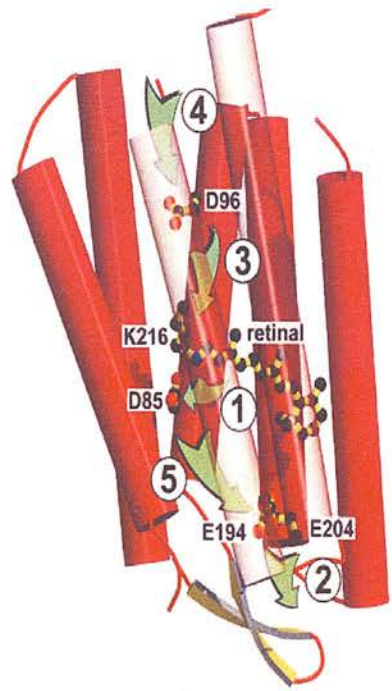


Figure 1.10 Bacteriorhodopsin with the steps in the proton transfer pathway shown. Key residues such as aspartic acid 96 and glutamic acids 194 and 204 are marked.

Kandt *et al* [17] used classical MD methods to simulate the diffusion of both the internal and external water molecules of the bacteriorhodopsin protein. Long simulation times, achievable due to the use of classical dynamics, allowed for the study of their distribution and showed good agreement with experimental work. Classical methods are reasonable to use with these types of models as there were no chemical reactions, and so there is no need to include the electronic interactions. Valuable

information on the preferred positions of water molecules (or H-bonding sites) can be obtained from such studies, giving a good insight into the proton pathway.

An attempt to model the water networks within bacteriorhodopsin has been reported that used a model of the entire protein embedded in a lipid bilayer [18]. A QM/MM approach was used due to the large size of the model (*ca.* 30000 atoms) with the QM used to model a small area of interest. The motivation for the study was to construct a working model of the system and then to add in both an Eigen and a Zundel complex to a hydrophilic pocket within the channel. The Eigen complex formed a stable solvated structure by hydrogen-bonding to amino acid residues within the pocket, as did the Zundel, but only with some restraints. The stability of these structures supports the theory of protonated networks within the protein framework.

A more focused investigation into the proposed mechanism of proton transport modelled the conformation change from the M to N states (during which the proton is transferred from Asp96 to the Schiff base).[19] A 1D H-bonded chain of water molecules extends from the Asp96 to the Schiff base, even through hydrophobic channel regions. A four water chain was used as a starting point and PT was observed. The addition of a fifth water molecule resulted in a more dynamical system with return events and clustering. This clustering can trap the proton, as seen in smooth walled channel simulations, in an Eigen type state which can halt the transfer process.

Much more extensive work is needed in order to provide a definitive answer for the mechanism of PT through bacteriorhodopsin, and it is testament to the possible applications of these findings that such work continues.

1.3.4. Cytochrome C oxidase

Cytochrome C oxidase (CcO) is the terminal enzyme in the process of cellular respiration and is used to both catalyse the reduction of dioxygen to water and also to translocate four excess protons to generate an electrochemical gradient which is subsequently used in ATP synthesis [1] (see Equation 1.1).

$$4e^- + 8H^+(in) + O_2 \rightarrow 2H_2O + 4H^+(out) \quad \text{.... (1.1)}$$

There is more than one possible PT pathway through the cytochrome but it is the so-called D-pathway, beginning with an aspartic acid residue and ending with a glutamic acid, which is thought to provide most of the eight protons required for the reaction *via* a hydrogen-bonded chain of water molecules. X-ray crystallographic structures have shown a series of water molecules in the channel which would support this.[58] Extensive simulation work has therefore been carried out on this pathway in order to define the exact mechanism of PT and to track the positions of the water molecules of the chain.

R. Cukier [23] performed a QM/MM MD study of CcO in order to observe the water networks formed over the duration of the simulation. Starting with a model of CcO with 12 internal waters, he added a further 500 to saturate the channel and then allowed the MD to run for a long period of time while looking for persistent water networks within the structure. A cluster of two water molecules was found around the glutamic acid residue of the D-pathway. An excess proton was added to the cluster to investigate the preferred state of the glutamic acid. Small thermal fluctuations in the cluster and surrounding solvating species were enough to change the stable protonated state within a few femtoseconds. The D-pathway was also studied and rather than one continuous chain of water molecules from Aspartic Acid (Asp) to Glutamic Acid (Glu), two chains were found to exist that span from Asp to Asparagine (Asn) and then Asn to Glu. These results agreed well with crystallographic data at the time.

Xu *et al* [22, 59] focused on the D-pathway only and were able to successfully model PT along a chain of waters. The exact behaviour of the proton was shown to depend on the protonation state of the Glu residue – if it is deprotonated the proton can travel the entire length of the chain *via* a Grotthuss type mechanism, if protonated it became trapped halfway.[59] Subsequent modelling of the proton trap [22] showed that the proton was delocalised over four water molecules in total, but would reside preferentially in an Eigen-like formation, with a hydronium ion at the centre. This agrees with previous findings stated from synthetic and smooth walled channels.

1.4. Transmembrane proteins: structure and environment

The proteins targeted for this research are integral or trans-membrane proteins (TMP) that span the entire lipid membrane, thus creating a pore through which ions can pass. Lipid membranes have a bi-layer structure with a hydrophilic outer surface and an internal hydrophobic region; the pore is created by the packing together of a number of individually stable alpha helices [see Figure 1.11(a)].

A class of proteins that has been adopted as a general model for TMPs is one based on the structure of bacteriorhodopsin [see Figure 1.11(b)]. This comprises seven alpha helix bundles between which reside central water wires to facilitate PT. In modelling work it is common practice for the entire protein and membrane to be included, with such a large model necessitating the use of classical simulation techniques. However under these conditions it is difficult to focus on the PT steps. The aim of this work is to move forward from this approach and use simpler models coupled with higher level computational techniques to probe the PT process more directly.

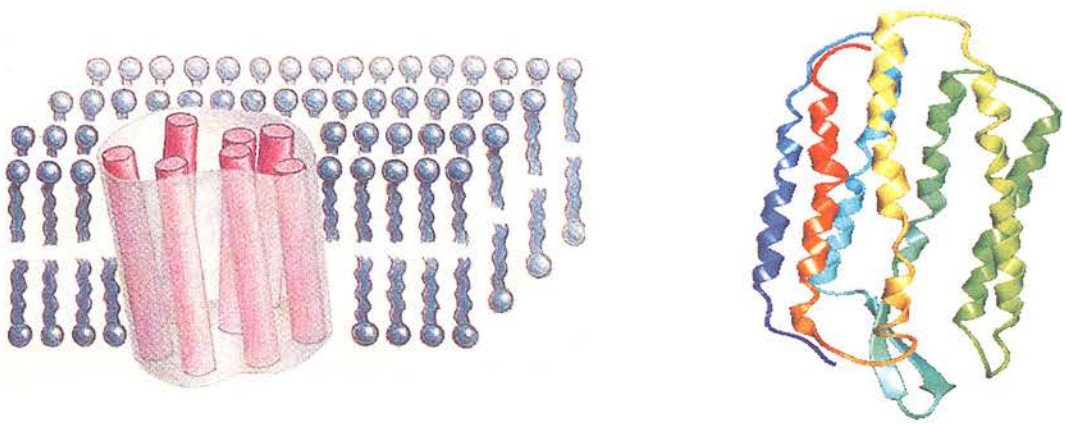


Figure 1.11 (a) Schematic of a trans-membrane protein and (b) The heptameric helical bundle of bacteriorhodopsin used as a model for trans-membrane protein.

1.5 Model systems

A further model system which has been shown to support proton transport is the β-helical polyglycine analogue of the gramcidin A channel with a water wire inserted into the internal cavity of the helix.[60] This model was small enough to study using quantum mechanical methods. In fact this model-concept is very similar to the work presented here, but we take a polyglycine α-helix expressed under periodic boundary conditions (PBCs) as our base model, which mimics a tertiary structure much more commonly found in TMPs.

1.5.1 The poly-glycine model system

The proteins targeted for this research (TMPs) span the entire lipid membrane, thus creating a pore through which ions can pass. The simplest possible model for PT in complex TMPs is a single helix and water wire (chain of water molecules plus an extra proton). An alpha helix is a series of amino acids ($NH_2CHRCOOH$) linked *via* peptide bonds to form a spring-like structure. The side chain R-groups point almost perpendicular to the direction of the helix (see Figure 1.12). The structure of the helix is retained almost regardless of what amino acids are used. Thus it is possible to greatly vary the sequence of amino acids in the helix whilst still retaining a shape compatible with life. Starting with a simple non- interacting amino acid chain such as polyglycine (R=H), we can build an understanding of the basic PT process. Replacing glycine with a polar or charged amino acid such as aspartic acid ($R=CH_2COOH$) will introduce more interaction between the helix and the water wire. Further and more complex models of the helix can then be obtained using different substitution patterns, with, for example, every fourth amino acid in a polyglycine chain altered to study the inward facing R-group interactions. Eventually the helix can be tuned to reflect the actual polypeptide sequence of a particular PT pathway, or important sections of it.

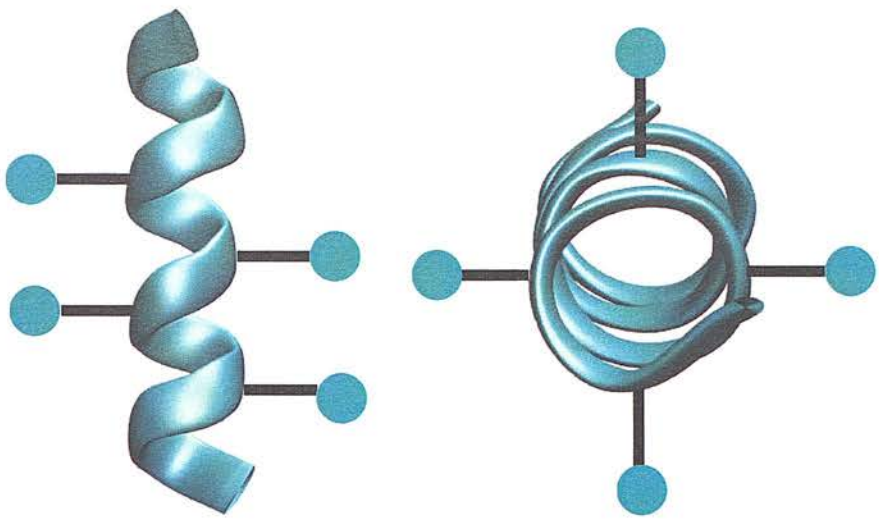
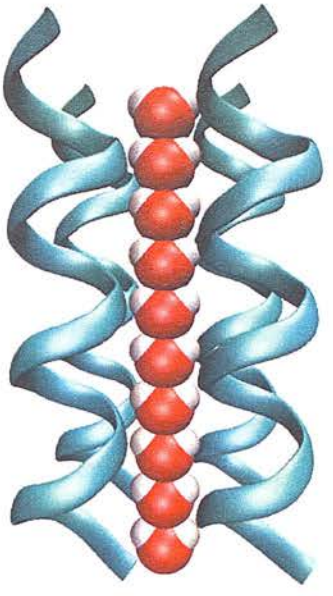


Figure 1.12 The alpha helix structure with the side chains represented with filled circles and backbone as a ribbon

To take a specific example of an alpha-helical domain, the simplest and most common structural motif consists of a bundle of four almost parallel helices.[61] We can mimic this with a model constructed from an orthorhombic box with lattice vectors $a = c$; the alpha-helices and water chain will lie along the b-axis [see Figure 1.13(a)]. PBCs can then be applied to build an infinite array of continuous channels, removing the requirement to build in a water reservoir at the entrance and exit sites of the channel [see Figure 1.13(b)]. The inter-helical distance (and thus the diameter of the central cavity) is varied directly by the a,c lattice vectors. Due to the demands of the PBCs the length of the b-vector relates to a single alpha-helical pitch, which also happily corresponds to the typical depth of a trans-membrane protein. By making use of PBCs in this way, we can therefore utilise just one alpha-helix and one water environment to build our channel environment. The downside is, of course, that we must accept that the infinite arrays are composed of identical copies of these building blocks, but the tremendous upside is that this model-system comprises less than 150 atoms, meaning that full quantum mechanical simulations are a realistic option, thus allowing the bond making and breaking steps to be treated appropriately, and MD simulations of the order of tens of picoseconds can be realistically harvested, thereby minimising the

problems associated with rare-event sampling. In fact, to the best of our knowledge, these simulations are the first report of an alpha-helical channel/water-wire system performed entirely at the quantum mechanical level.

a)



(b)

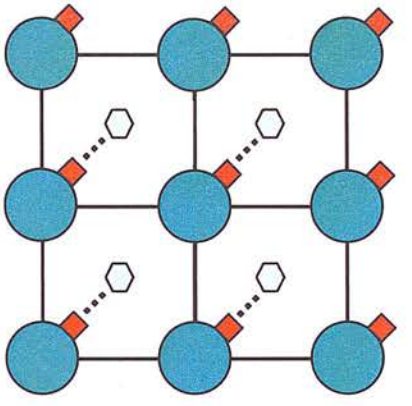


Figure 1.13 (a) Four helical bundle with water wire from MD simulation (b) Representation of the full PBC model used to mimic a tetrameric protein where the large circles are helices, the small hexagons water wires and the squares represent the inward facing side chain R-groups

A potential criticism can, however, be levied at the use of *ab initio* MD calculations when applied to study PT reactions. The many-electron Hamiltonian is approximated by density functional theory (DFT, discussed in the next chapter), and it is well known that current exchange-correlation functionals lack the non-local contributions essential for the accurate modelling of a bond stretched to its breaking point, resulting typically in an underestimation of PT barrier heights. A further problem which arises in these simulations is the neglect of zero-point energy contributions (ZPC) in classical dynamics, which will result in an over-estimation of the barrier height. There are therefore two competing deficiencies in our simulations, which will partially cancel in a very fortuitous manner.[62] The pure DFT functional will underestimate the PT barrier, but the neglect of the ZPC will work against this to our favour.

It must also of course be stressed that the structures of actual trans-membrane channels are significantly more complex than our model allows. Our work does however help to narrow the gap between the constrained environment models previously reported and the fully atomistic protein studies that can only be modelled using lower-level computational techniques.

The computational methods used throughout the thesis are discussed in more detail in Chapter 2. Results and analysis from the three model systems (poly-glycine, poly-glycine-serine and poly-glycine-aspartic acid) are presented in Chapters 3-5 and final conclusions in Chapter 6.

1.6. Bibliography

1. Wraight, C.A., *Chance and design--Proton transfer in water, channels and bioenergetic proteins.* Biochimica et Biophysica Acta (BBA) - Bioenergetics, 2006. **1757**: p. 886-912.

2. Decoursey, T.E., *Voltage-Gated Proton Channels and Other Proton Transfer Pathways.* Physiological Review, 2003. **83**: p. 475-579.

3. Swanson, J.M.J., et al., *Proton Solvation and Transport in Aqueous and Biomolecular Systems: Insights from Computer Simulations.* Journal of Physical Chemistry B, 2007. **111**: p. 4300-4314.

4. Subramaniam, S., T. Hirai, and R. Henderson, *From structure to mechanism: electron crystallographic studies of bacteriorhodopsin.* Philosophical Transactions of the Royal Society A, 2002. **360**: p. 859-874.

5. Nishimura, K., et al., *The Closed State of a H+ Channel Helical Bundle Combining Precise Orientational and Distance Restraints from Solid State NMR.* Biochemistry, 2002. **41**: p. 13170-13177.

6. Braun, T.F., et al., *Arrangement of Core Membrane Segments in the MotA/MotB Proton-Channel Complex of Escherichia coli.* Biochemistry, 2004. **43**: p. 35-45.

7. Preston, G.M., et al., *Appearance of Water Channels in Xenopus Oocytes Expressing Red Cell CHIP28 Protein.* Science, 1992. **256**: p. 385-387.

8. Findlay, J. and E. Eliopoulos, *Three-dimensional modelling of G protein-linked receptors.* Trends in Pharmacological Sciences, 1990. **11**: p. 492-499.

9. Pardo, L.J., et al., *On the use of the transmembrane domain of bacteriorhodopsin as a template for modeling the three-dimensional structure of guanine nucleotide-binding regulatory protein-coupled receptors.* Proceedings of the National Academy of Sciences of the United States of America, 1992. **89**: p. 4009-4012.

10. Patny, A., P.V. Desai, and M. Avery, *Homology Modelling of G-Protein Coupled Receptors and Implications in Drug Design.* Current Medicinal Chemistry, 2006. **13**: p. 1667-1691.

11. Luecke, H., et al., *Structure of bacteriorhodopsin at 1.55 Å resolution.* Journal of Molecular Biology, 1999. **291**: p. 899-911.

12. Schmitt, U.W. and G.A. Voth, *Multistate Empirical Valence Bond Model for Proton Transport in Water.* Journal of Physical Chemistry B, 1998. **102**: p. 5547-5551.

13. Day, T.J.F., et al., *A second generation multistate empirical valence bond model for proton transport in aqueous systems.* The Journal of Chemical Physics, 2002. **117**: p. 5839-5849.

14. Wu, Y., et al., *An Improved Multistate Empirical Valence Bond Model for Aqueous Proton Solvation and Transportation.* The Journal of Physical Chemistry B, 2008. **112**: p. 467.

15. Riccardi, D., et al., *Development of Effective Quantum Mechanical/Molecular Mechanical (QM/MM) Methods for Complex Biological Processes.* The Journal of Physical Chemistry B, 2006. **110**(13): p. 6458-6469.

16. Ignacio Fdez, G., et al., *A QM/MM study of proton transport pathways in a [NiFe] hydrogenase.* Proteins: Structure, Function, and Bioinformatics, 2008. **73**(1): p. 195-203.

17. Kandt, C., K. Gewert, and J. Schlitter, *Water dynamics simulation as a tool for probing proton transfer pathways in a heptahelical membrane protein.* Proteins: Structure, Function, and Bioinformatics, 2005. **58**: p. 528-537.

18. Rousseau, R.V.K., et al., *Assigning Protonation Patterns in Water Networks in Bacteriorhodopsin Based on Computed IR Spectra.* Angewandte Chemie International Edition, 2004. **43**: p. 4804-4807.

19. Lee, Y.-S. and M. Krauss, *Reversible proton transfer dynamics in bacteriorhodopsin.* Journal of Molecular Structure, 2004. **700**: p. 243-246.

20. Smondyrev, A.M. and G.A. Voth, *Molecular Dynamics Simulation of Proton Transport through the Influenza A Virus M2 Channel.* Biophysical Journal, 2002. **83**: p. 1987-1996.

21. Chen, H., Y. Wu, and G.A. Voth, *Proton Transport Behavior through the Influenza A M2 Channel: Insights from Molecular Simulation.* Biophysical Journal, 2007. **93**: p. 3470-3479.

22. Xu, J., et al., *Storage of an Excess Proton in the Hydrogen-Bonded Network of the D-Pathway of Cytochrome c Oxidase: Identification of a Protonated Water Cluster.* Journal of the American Chemical Society, 2007. **129**: p. 2910-2913.

23. Cukier, R.I., *Quantum molecular dynamics simulation of proton transfer in cytochrome c oxidase.* Biochimica et Biophysica Acta (BBA) - Bioenergetics, 2004. **1656**: p. 189-202.

24. Wikström, M., M.I. Verkhovsky, and G. Hummer, *Water-gated mechanism of proton translocation by cytochrome c oxidase.* Biochimica et Biophysica Acta (BBA) - Bioenergetics, 2003. **1604**: p. 61-65.

25. Zheng, X., et al., *Computer simulation of water in cytochrome c oxidase.* Biochimica et Biophysica Acta (BBA) - Bioenergetics, 2003. **1557**: p. 99-107.

26. Lear, J.D., Z.R. Wasserman, and W.F. DeGrado, *Synthetic Amphiphilic Peptide Models for Protein Ion Channels.* Science, 1988. **240**: p. 1177-1181.

27. Wu, Y., B. Ilan, and G.A. Voth, *Charge Delocalization in Proton Channels, II: The Synthetic LS2 Channel and Proton Selectivity.* Biophysical Journal, 2007. **92**: p. 61-69.

28. Wu, Y. and G.A. Voth, *A Computer Simulation Study of the Hydrated Proton in a Synthetic Proton Channel.* Biophysical Journal, 2003. **85**: p. 864-875.

29. Chen, H., et al., *Charge Delocalization in Proton Channels, I: The Aquaporin Channels and Proton Blockage.* Biophysical Journal, 2007. **92**(1): p. 46-60.

30. Agmon, N., *The Grotthuss mechanism.* Chemical Physics Letters, 1995. **244**: p. 456-462.

31. Schmitt, U.W. and G.A. Voth, *The computer simulation of proton transport in water.* The Journal of Chemical Physics, 1999. **111**: p. 9361-9381.

32. Lapid, H., et al., *A bond-order analysis of the mechanism for hydrated proton mobility in liquid water.* The Journal of Chemical Physics, 2005. **122**: 014506.

33. Tuckerman, M.E., et al., *On the Quantum Nature of the Shared Proton in Hydrogen Bonds.* Science, 1997. **275**: p. 817-820.

34. Marx, D., et al., *The nature of the hydrated excess proton in water.* Nature, 1999. **397**: p. 601-604.

35. Marx, D., M. Tuckerman, E. , and M. Parrinello, *Solvated excess protons in water: quantum effects on the hydration structure.* Journal of Physics: Condensed Matter, 2000: p. A153.

36. Day, T.J.F., U.W. Schmitt, and G.A. Voth, *The Mechanism of Hydrated Proton Transport in Water.* The Journal of the American Chemical Society, 2000. **122**: p. 12027-12028.

37. Voth, G.A., *Computer Simulation of Proton Solvation and Transport in Aqueous and Biomolecular Systems.* Accounts of Chemical Research, 2006. **39**: p. 143-150.

38. Berkelbach, T.C., H.-S. Lee, and M.E. Tuckerman, *Concerted Hydrogen-Bond Dynamics in the Transport Mechanism of the Hydrated Proton: A First-Principles Molecular Dynamics Study.* Physical Review Letters, 2009. **103**(23): 238302.

39. Pomes, R. and B. Roux, *Theoretical Study of H+ Translocation along a Model Proton Wire.* Journal of Physical Chemistry, 1996. **100**: p. 2519-2527.

40. Pomes, R. and B. Roux, *Free Energy Profiles for H+ Conduction along Hydrogen-Bonded Chains of Water Molecules.* Biophysical Journal, 1998. **75**: p. 33-40.

41. Brewer, M.L., U.W. Schmitt, and G.A. Voth, *The Formation and Dynamics of Proton Wires in Channel Environments.* Biophysical Journal, 2001. **80**: p. 1691-1702.

42. Dellago, C. and G. Hummer, *Kinetics and Mechanism of Proton Transport across Membrane Nanopores.* Physical Review Letters, 2006. **97**: 245901.

43. Dellago, C., M.M. Naor, and G. Hummer, *Proton Transport through Water-Filled Carbon Nanotubes.* Physical Review Letters, 2003. **90**: 105902.

44. Hummer, G., *Water, proton, and ion transport: from nanotubes to proteins.* Molecular Physics, 2007. **105**: p. 201-207.

45. Nagle, J.F. and H.J. Morowitz, *Molecular Mechanisms for Proton Transport in Membranes.* Proceedings of the National Academy of Sciences of the United States of America, 1978. **75**: p. 298-302.

46. Pearlman, D.A., et al., *AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules.* Computer Physics Communications, 1995. **91**: p. 1-41.

47. MacKerel Jr, A.D., et al., *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program*. 1998, John Wiley & Sons: Chichester. p. 271-277.

48. Wu, Y., et al., *An Improved Multistate Empirical Valence Bond Model for Aqueous Proton Solvation and Transportation.* The Journal of Physical Chemistry B, 2007. **112**: p. 467.

49. Akeson, M. and D.W. Deamer, *Proton conductance by the gramicidin water wire: A Model for proton conductance in the F1F0 ATPases.* Biophysical Journal, 1991. **60**: p. 101-109.

50. Levitt, D.G., S.R. Elias, and J.M. Hautman, *Number of water molecules coupled to the transport of sodium, potassium and hydrogen ions via gramicidin, nonactin or valinomycin.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 1978. **512**: p. 436-451.

51. Dani, J.A. and D.G. Levitt, *Water transport and ion-water interaction in the Gramicidin channel.* Biophysical Journal, 1981. **35**: p. 501-508.

52. Sagnella, D.E. and G.A. Voth, *Structure and dynamics of hydronium in the ion channel Gramicidin A.* Biophysical Journal., 1996. **70**: p. 2043-2051.

53. Pomes, R. and B. Roux, *Structure and dynamics of a proton wire: a theoretical study of H+ translocation along the single-file water chain in the Gramicidin A channel.* Biophysical Journal, 1996. **71**: p. 19-39.

54. Pomes, R. and B. Roux, *Molecular Mechanism of H+ Conduction in the Single-File Water Chain of the Gramicidin Channel.* Biophysical Journal, 2002. **82**: p. 2304-2316.

55. Helenius, A., *Unpacking the incoming influenza virus.* Cell, 1992. **69**: p. 577-578.

56. Wang, C., R.A. Lamb, and L.H. Pinto, *Activation of the M2 ion channel of influenza virus: a role for the transmembrane domain histidine residue.* Biophysical Journal, 1995. **69**: p. 1363-1371.

57. Shuck, K., R.A. Lamb, and L.H. Pinto, *Analysis of the Pore Structure of the Influenza A Virus M2 Ion Channel by the Substituted-Cysteine Accessibility Method.* Journal of Virology., 2000. **74**: p. 7755-7761.

58. Qin, L., et al., *Identification of conserved lipid/detergent-binding sites in a high-resolution structure of the membrane protein Cytochrome C Oxidase.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**: p. 16117-16122.

59. Xu, J. and G.A. Voth, *Chemical Theory and Computation Special Feature: Computer simulation of explicit proton translocation in Cytochrome C Oxidase: The D-pathway.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**: p. 6795-6800.

60. Sagnella, D.E., K. Laasonen, and M.L. Klein, *Ab initio Molecular Dynamics Study of Proton Transfer in a Polyglycine Analog of the Ion Channel Gramicidin A.* Biophysical Journal, 1996. **71**: p. 1172-1178.

61. Branden, C. and J. Tooze, *Introduction to Protein Structure*. 2nd ed. 1999, New York: Garland Publishing, Inc.

62. Durlak, P., et al., *Car-Parrinello and path integral molecular dynamics study of the hydrogen bond in the chloroacetic acid dimer system.* The Journal of Chemical Physics, 2007. **127**: p. 64304-64308.

Chapter 2:

Simulation Methods

A description of density functional theory and its application in molecular modeling programs

2.1 Introduction

Computational simulation and modelling is rapidly becoming one of the most sought after resources in chemistry and has found application in many areas, with perhaps the most well known being to aid drug discovery for the pharmaceutical industry. In this application it can dramatically improve both the number of potential drug leads and also reduce the timescale of bringing a new drug to market. It is being used more widely in the chemical industry, particularly in the area of nanotechnology. It has also found use in academia to probe reactions and systems to provide answers to questions that were previously out of reach.

Calculating the macroscopic properties of a material is a complex problem and at present there are two different approaches used. The first, known as classical mechanics, uses a force field, which is a potential describing all the interactions between atoms within the system. The success of the force field is reliant on the fitting of empirical potentials (e.g. Lennard-Jones potentials) to a set of parameters derived from other methods. The second is based on first principles and the de Broglie wave-particle duality of electrons, which allows for them to be described using a wavefunction. The square of this wavefunction determines the probability of an electron being at a certain position at a certain time. As first-principles methods deal explicitly with electron probabilities, they are capable of modelling chemical reactions, where bonds between atoms can be formed and broken. It is this level of detail that is required to model proton transfer events; to this end this chapter is devoted to a basic discussion of first-principles simulation methods. [1-4]

2.2. The Schrödinger Equation

Using de Broglie wave-particle duality Schrödinger derived his wave equation whereby an operator, $\hat{H}$, acts upon the wavefunction to give the energy of the system multiplied by the wavefunction (Equation 2.1).

$$\hat{H}\Psi(r,R) = E\Psi(r,R) \qquad \text{.... (2.1)}$$

The Hamiltonian operator, $\hat{H}$, describes five interactions of the system:

1. the kinetic energy of the electrons;
2. the kinetic energy of the nuclei;
3. the potential energy of nuclear-nuclear repulsion;
4. the potential energy of electron-electron repulsion; and
5. the potential energy of nuclear-electronic attraction.

In full quantum mechanical terms we have the following:

$$\hat{H} = -\sum_i \frac{\hbar^2}{2m_e}\nabla_i^2 - \sum_k \frac{\hbar^2}{2m_n}\nabla_k^2 - \sum_i\sum_k \frac{e^2 Z_k}{4\pi\varepsilon_0 r_{ik}} - \sum_{i<j} \frac{e^2}{4\pi\varepsilon_0 r_{ij}} + \sum_{k<l} \frac{e^2 Z_k Z_l}{4\pi\varepsilon_0 r_{kl}} \qquad \text{.... (2.2)}$$

where i and j run over all the electrons of the system, k and l over all the nuclei, $\hbar$ is Planck's constant divided by 2π, m_e is the mass of an electron, m_n is the mass of a nucleus, e is the charge of an electron, ∇ is the Laplacian operator (which introduces x, y, and z dimensionality), ε_0 is the vacuum permittivity, Z is the atomic number of an atom and r_{ab} is the distance between particles a and b.

If all interactions of the system are known exactly then this equation can be solved and all properties of the system are known. In practice this can only be achieved for one particle systems such as the particle in a box, the harmonic oscillator and the hydrogen atom. For all other systems an exact solution is impossible; we must invoke some approximations.

2.3. Approximations

The Born-Oppenheimer (BO) approximation allows for the decoupling of electronic and nuclear motion, based on the observation that electrons travel orders of

magnitude faster than nuclei. The mass of the smallest nuclei, the proton, is 1840 times heavier than the resting mass of an electron. In practice this means that the electrons will be seen to adjust instantaneously to any movement of the nuclei. The wavefunction can thus be separated into ionic and electronic parts:

$$\Psi(r,R,t) = \theta(R,t).\Psi_e(r,\{R\}), \qquad \text{.... (2.3)}$$

where θ(R, t) is the ionic part of the wavefunction, which depends only on the ionic positions and time, and Ψ_e(r, R) is the electronic part of the wavefunction, depending on the positions of the electrons and with the ionic positions as parameters. We can therefore assume, from the electronic perspective that the nuclei are stationary. From this standpoint two of the five terms in Equation 2.2 are greatly simplified: the second term (the kinetic energy of the nuclei) is set to zero and term three (the nuclear-nuclear repulsion term) becomes a constant that can be described classically by Coulomb's law. Thus the nuclear terms can be taken out of the Hamiltonian, leaving only the electronic elements to be considered further.

The use of this approximation allows for the simplification of the Schrödinger equation to a form which is solvable, but only for hydrogenic systems, *i.e.* those with only one electron such as H_2^+, H and He^+. Moving to more complicated many electron systems means that the electron-electron repulsion term in the Hamiltonian cannot be solved. Overcoming this problem is the greatest challenge in molecular modelling. Two factors contribute: (i) Electron correlation – whereby the motion of one electron is directly affected by the path of those around it and (ii) Electron exchange – electrons with differing spin can pair up and occupy the same space or orbital. The ways in which these two issues are addressed gives rise to the level of theory used for the calculation.

2.4. Hartree-Fock Theory

The problem of electron repulsion cannot be ignored. At a distance of 1Å the interaction energy is approximately 14eV [5] – which if ignored would lead to

completely inaccurate results. Hartree developed a method whereby repulsion is accounted for but correlation and exchange ignored. Imagine one electron moving around a nucleus. The potential generated due to the interaction of an electron with a nucleus will fluctuate due to the movement of the electron. If more electrons are present they too will contribute to these fluctuations. At a particular point in space, however, there will be an average potential, $V_m(r)$. Hartree devised a method which allows each electron to move in a constant field, $V_{eff}(r)$ generated by the other electrons present, similar to the average potential. This potential does not fluctuate – but the electrons feel each others presence, if only in an average way.

The second problem is electron exchange. Electrons are completely indistinguishable from each other, but they cannot occupy the same space, *i.e.* they cannot have the same four quantum numbers. Their wavefunction is dependent on whether they are 'spin up' or 'spin down' and if any two electrons are to be interchanged, their spin, and therefore their wavefunction sign must be inverted. Fock added exchange to Hartree theory with the inclusion of an asymmetric wavefunction,

$$\Psi_a \uparrow (r_1)\Psi_b \uparrow (r_2) - \Psi_a \uparrow (r_2)\Psi_b \uparrow (r_1) \qquad \text{.... (2.4)}$$

where $\uparrow$ represents the spin of the electron. This wavefunction does not allow for the electrons to occupy the same space, if $r_1 = r_2$ then the wavefunction, and probability are reduced to zero.

The effective field of Hartree, which essentially removes direct electron-electron interaction, along with the asymmetric wavefunction allows for the decoupling of the electrons. The Schrödinger equation can now be solved one electron at a time; the resulting N equations are then summed to generate one overall solution. This combined Hartree-Fock (HF) theory can account for up to 99% of the real energy of a system, with the missing energy attributable to the missing correlation energy. There are improvements to HF which use higher order functions and attempt to approximate the missing (correlation) energy – these include the Møller-Plesset series, configuration interaction and also couple cluster levels of theory.

2.5. Density Functional Theory

Hartree-Fock theory is based on $4N$ variables where N is the number of electrons – three spatial variables and one exchange *i.e.* spin up or down. An alternative method based on only the 3 spatial variables has been developed, which has at its heart the calculation of the experimentally verifiable electron density. This method is called density functional theory (DFT). It is based on two main theorems: (i) The total energy of the system can be directly related to its electron density by a unique functional and (ii) The minimum value of the functional is equal to the ground state energy of the system in that position.[6]

In DFT the Hamiltonian operator is again formalised for a non-interacting system of electrons giving N one electron equations (equivalent to individual Schrödinger equations in HF theory). These are termed the 'Kohn-Sham' equations, [7]

$$\left[\frac{-\hbar^2}{2m}\nabla^2 + V_{ion}(r) + V_H(r) + V_{XC}(r)\right]\Psi_i(r) = \varepsilon_i \Psi_i(r), \qquad \text{.... (2.5)}$$

where Ψ_i is the wavefunction of the ionic state *i*, ε_i is the Kohn-Sham eigenvalue, V_H is the Hartree potential of the electrons and V_{XC} is the exchange-correlation potential.

Hohenburg and Kohn stated that the energy is related to the density by a unique functional. However this functional is not exactly known, representing the major drawback to DFT. The DFT method approximates both exchange and correlation, whereas HF theory approximates correlation but gets exchange exactly right. DFT functionals essentially model 'electron holes' whereby electrons can dig themselves into a 'well' which prevents other electrons from occupying the same space or even getting too close. These approximations also only work well locally, so there is no way to account for long range or dispersive forces in DFT.

As a result a series of functionals have been developed, based on different assumptions. The most basic is the local density approximation (LDA), which is designed for metallic systems where the electron density does not fluctuate greatly. An improvement on this from a chemical perspective is the generalised gradient

approximation (GGA) which includes information about the gradient of changing electron density within a molecular framework. Examples of these functionals include PBE of Perdew, Burke, and Ernzerhof [8], PW91 of Perdew and Wang [9] and BLYP, composed of the Becke exchange functional [10] (B) and the correlation functional of Lee, Yang and Parr [11] (LYP). Which of these functionals is used depends very much on the system to be studied. Part of the work presented in Chapter 3 explores the importance and implications of functional choice to model accuracy. As a result of this work the BLYP functional has been used throughout.

2.6. Basis Sets

The wavefunction of the Hartree-Fock or Kohn-Sham equations is approximated using a series of basis functions (or atomic orbitals) known as a basis set. This method is based on the linear combination of atomic orbitals (LCAO), whereby the functions will combine linearly to describe the electron density of the entire system. A series of functions (Equation 2.6) are chosen and weighted using a weighting coefficient, *c*, in order to generate the best wavefunction which will, *via* the variational theorem, give the lowest energy configuration for that arrangement of atoms.

$$\psi_i(r) = \sum_{\alpha} c_{i\alpha}\, \phi_\alpha(r) \qquad \text{.... (2.6)}$$

There are two types of basis sets used: (i) localised (where the electrons are assumed to reside close to the atomic centres) and (ii) delocalised (where electrons are modelled as nearly-free particles).

2.6.1. Localised Basis Sets

Localised basis sets are based on atomic radial functions (ARF) of atoms, the shapes of which are approximated using Gaussian functions (see Equation 2.

$$g(r) = \left(\frac{2\alpha}{\pi}\right)^{0.75} e^{(\alpha . r^{-2})} \qquad \text{.... (2.7)}$$

A single Gaussian cannot be used as they decay too rapidly at the edges and cannot accurately describe the electron density close to the nucleus. This problem is overcome by using a number of Gaussians with varying 'spread', depending on α. These are summed together to give an estimate of the ARF (see Figure 2.1). The weighting of each Gaussian is iteratively improved until the sum of all gives the best fit (i.e. the best wavefunction, which in turn generates the lowest energy). There are many levels of basis set which increase in accuracy as the number of Gaussians functions used to describe the orbitals is increased.

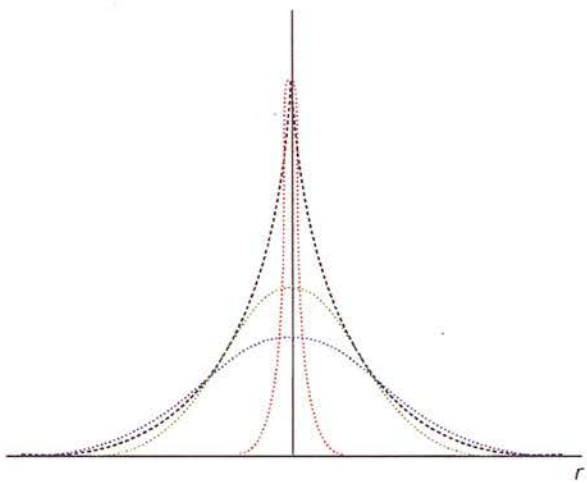


Figure 2.1. The three coloured Gaussian functions are summed together to construct the hydrogen 1s orbital, shown as a solid black line

The valence electrons are those which are more chemically relevant and thus it makes more sense to use more computational effort to model them versus the core electrons. In order to accomplish this, basis sets are normally split into two sets of functions, one set to describe the core electrons and a second used to represent the valence electrons. Pople [12, 13] developed these split-valence basis sets - such as 3-21G. This example is a minimal split valence set which uses 3 Gaussians to describe the core and then 2 further blocks of 2 and 1 to describe the valence electrons. Increasing quality means adding more functions to the valence and core blocks, *e.g.* 6-311G.

Further improvements can be made with the introduction of additional functions to include polarisation and diffusive effects. The polarisation functions enable the electrons to occupy higher quantum states than their ground, such as *p*-type orbitals for an *s*-electron. Diffuse functions are added to describe the electrons when they are far from the nucleus e.g. excited species and anions. The polarisation functions are indicated by an asterisk (6-31G*) and diffuse using a plus symbol (6-31+G).

2.6.2. Delocalised Basis Sets

The second type of basis set is a delocalised or plane wave basis set. The electrons are modelled as nearly free particles and described by the function,

$$\Psi_i(r) = \Sigma_k c_{ik} \exp(ik.r) \qquad \text{.... (2.8)}$$

where k is the wavevector of the de Broglie wave (reciprocal of the wavelength) and c_{ik} are the weighting coefficients that are varied until the ground state wavefunction and energy state are reached. Real systems contain electrons of greatly varying energies and thus a large number of functions of differing energy, k values, are required to model all the electrons of a system. The quality of these basis sets is determined by E_{cut} ranging from 250eV (poor basis set) to 600eV (very good basis set with typically many thousands of plane waves).

There are many advantages to this method. The first is the continuous nature of plane wave basis sets – they can be systematically improved simply by increasing the E_{cut} value to include higher energy functions. They are relatively simple and easy to compute and are periodic, which allows for the use of periodic boundary conditions (PBCs, see next section). Similar to the Gaussian approach, modelling the core electrons is computationally expensive. But they cannot simply be removed from the model. A pseudopotential method was introduced which replaces the core electrons with a smooth effective potential (see Figure 2.2). This potential allows the effect of the core to be felt, but softens any interference with valence electrons and reduces considerably the necessary basis set quality. Examples of pseudopotentials include Vanderbilt ultrasoft potentials [14] and also Hamann's norm-conserving potentials.[15]

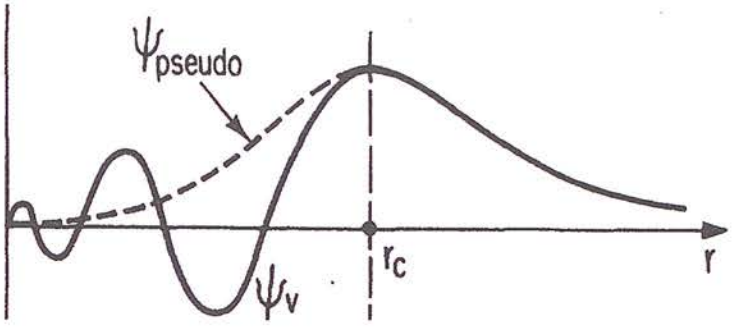


Figure 2.2. A schematic of the pseudopotential method. The all electron wavefunction is shown by a solid black line. The pseudopotential is used to smooth over the wavefunction up to an E_{cut} value (r_c) when the original potential is restored

2.6.3 Combining basis sets

The computational method used in this thesis employs a mixed-basis set approach. The principle of this method is that the orbitals are represented using standard Gaussian functions, while the density defined within this space is represented by delocalised plane waves. Thus electrons are now assumed to spend their time located on the atoms, rather than wandering as almost free-particles throughout the system. This represents a clear potential for computational speed-up over the standard plane-wave basis sets approach. In fact, the method can provide an almost linear

scaling with system size and as such allows for increased number of atoms and longer molecular dynamics (MD) production trajectories to be obtained. And as with standard Gaussian basis sets, those orbitals that are sufficiently far apart can be ignored. The code uses neighbour lists, like classical MD, to define these non-important interactions. This is a technical point, but one which adds further to the potential speed up in calculation time offered by this dual-basis set approach.

2.7 Periodic Boundary Conditions

Even with the advances in computational power the system size amenable to simulation is still relatively small. In order to enable larger, more relevant structures or

bulk fluids to be studied periodic boundary conditions are employed. A two or three dimensional box containing the basic building blocks of the model is taken and repeated to infinity to create an array which is used to simulate the bulk material. There are a number of cell shapes used, including truncated octahedron and hexagonal prism, but in this work a simple cubic cell is used. The cells are chosen in order that they can pack together exactly, to avoid any spaces or artificial surfaces/boundaries. The mainstay of this method is that the cells have soft boundaries so that any escaping particles will reappear from the opposite cell side. There are of course issues with this approach, in that the cell will 'feel' its neighbours and this may lead to artificial interactions and long range ordering of the bulk.

2.8 Geometry Optimisation

Geometry optimisation is the process by which the lowest energy, and thus the most stable structure of a given molecule or system, is found. A starting structure is generated and the initial total energy of the system determined. A function of the energy versus reaction coordinates representing atom positions (in Cartesian or internal coordinates, *e.g.* bond lengths, angles) can be plotted (see Figure 2.3).

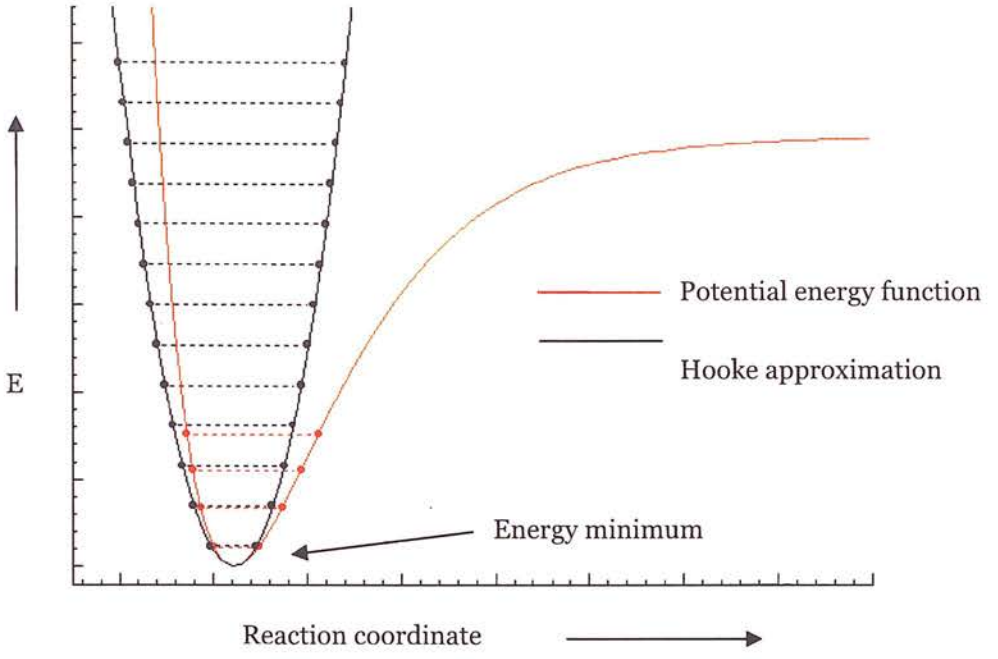


Figure 2.3. A plot of energy *vs.* reaction coordinate, *e.g.* bond length. The potential energy function is shown in red and a modelling approximation based on Hooke's Law shown in black. An energy minimum or stationary point is marked.

To a simple approximation Hooke's Law can be used to model the function in the form:

$$E = \frac{1}{2}kx^2 \qquad \text{.... (2.9)}$$

where x is the reaction coordinate of interest, k is the force constant holding the atoms together and E is the total energy of the system. The first derivative (*i.e.* the gradient – equation 2.10) of the function is then used to calculate the forces acting upon the atoms at this particular geometry. The atoms are then moved in order to minimise the energy.

$$\frac{dE}{dx} = kx \qquad \text{.... (2.10)}$$

If the system is at an energy minima then the first derivative will be equal to zero, *i.e.* the atoms will not want to move. The purpose, therefore, of geometry optimisation is to

reach a structure where the forces are driven to zero, corresponding to a stationary point on the PES (see Figure 2.3).

Optimisations can be performed solely using first derivatives, *e.g.* steepest descents methods or conjugate gradients. Alternatively, the second derivative (yielding the force constant or curvature of the PES) can also be taken into account (see Equation 2.11):

$$\frac{d^2E}{dx^2} = k \qquad \text{.... (2.11)}$$

In practice it is only the estimates of the second derivatives that are utilised, as analytic second derivative calculations are computationally very expensive. Examples of optimisation schemes based on this principle are the Berny [16] (implemented for localised basis set calculations) and the Broyden-Fletcher-Goldfarb-Shanno (BFGS) methods (used in plane wave calculations). [17-20]

To summarise, an optimised structure is obtained by calculating the forces and most probably also by estimating the force constants. The atoms are moved in an attempt to lower the energy and drive the forces to zero. The entire process is repeated in an iterative fashion until the energy (and the structure) does not change; the final structure can be compared to experimental data for verification.

2.9. Molecular Dynamics

Molecular dynamics (MD) simulations are used to calculate the 'real' dynamics of a system from which time averaged properties can be determined. The previously optimised structure is now 'given' a certain amount of energy (*i.e.* atomic velocities randomly assigned from a Maxwell-Boltzmann distribution for a particular temperature) and the atoms allowed to move within the constraint forces of the molecular framework. The position, *r*, of each atom at a particular time, *t*, is calculated based on its previous position. The expression for this can be written as a Taylor expansion:

$$r(t+\partial t) = r(t) + \partial t\, v(t) + \frac{1}{2}\partial t^2 a(t) + \frac{1}{6}\partial t^3 b(t) + ...etc \qquad .... (2.12)$$

This is an exact relation; hence if the position of an atom at time *t*, and all the higher derivatives of the position *r* are known, it is possible to calculate the atom position at any time in the future, $t + \partial t$.

The first derivative is the velocity, *v*, and is easily calculated. The second is the acceleration of the atom, *a*, which can be obtained from the force exerted on that atom from Newton's second law of motion:

$$F_i = m_i a_i \qquad .... (2.13)$$

where F_i is the force acting on atom *i* (derived from either an empirical potential or from a quantum mechanical simulation), and m_i is the mass of the atom. Unfortunately the higher derivatives (denoting the change of force with time etc.) are inaccessible. If however the time step ∂t is very small then the impact of the higher derivatives is minimal and the truncated equation holds true:

$$r(t+\partial t) = r(t) + \partial t v(t) + \frac{1}{2}\partial t^2 a(t) \qquad .... (2.14)$$

This underlines the importance of choosing the correct value of ∂t for the simulation. Known as the 'time step' in an MD simulation it must be small enough to ensure the above approximation, and yet not so small that the simulation progresses in frustratingly short steps. In practise the time step is chosen in order to adequately sample the highest energy or characteristic vibration of the system. This is frequently an O-H or N-H vibration and the timestep must be short enough for the vibration to be sampled once per 10-15 MD steps.

In practise there are two main types of ab initio MD methods, and the timestep chosen is largely dictated by which of these methods is used. The first is based on the decoupling of the nuclear and electronic motion, as per the BO approximation (to give BOMD). This decoupling of the electronic motion from the nuclear means that the

timestep chosen is dependent only on nuclear motion; it is therefore relatively large. Each MD cycle, however, contains a series of steps (including multiple wavefunction minimisations) to calculate the forces, move the nuclei accordingly and compute the electronic structure which is expensive computationally.

An alternative approach is Car-Parrinello MD[21] (CPMD). Here the electronic and nuclear dynamics are performed simultaneously – there is no decoupling of motion. The electronic structure calculated does not have to be the ground state for that configuration and this means that there is only one wavefunction optimisation step per MD cycle. This method does generate errors in the nuclear forces, but it has been shown that these are fortuitously cancelled out by equal and opposite errors in the electronic motion. [4] The timestep used must be significantly smaller than that of BOMD, however, as now the electronic as well as the nuclear motion must now be sampled, though the added computer time from this is balanced by the fewer wavefunction minimisations needed. A CPMD time step is typically of the order of one-tenth required for a BOMD simulation.

2.9.1. Integrating the equations of motion

A robust integration scheme is crucial for a successful MD simulation. All of the integration schemes used in MD calculations assume that the new positions and velocities can be regarded as a sum of the truncated Taylor expansions (Eqn. 2.14). The most widely used approach, and also the simplest, is the velocity-Verlet scheme. In practice the implementation of the algorithm is a three step procedure: (i) the positions at $r(t+\delta t)$ are calculated using eqn. 2.15 with velocities and accelerations at time t; (ii) the velocities at a half step, $t+\delta t/2$ are determined using eqn 2.16 while the new forces are computed from the potential using the atom positions at $t+\delta t$ to give the acceleration; (iii) finally the new forces are computed from the potential using the atom positions at $t+\delta t$ to give the acceleration (Eqn 2.17).

$$r(t+\partial t) = r(t) + \partial t v(t) + \frac{1}{2}\partial t^2 a(t) \qquad \text{.... (2.14)}$$

$$v(t+\partial t) = v(t) + \frac{1}{2}\partial t[a(t) + a(t+\partial t)] \quad \text{.... (2.15)}$$

$$v(t+\frac{1}{2}\partial t) = v(t) + \frac{1}{2}\partial t a(t) \quad \text{.... (2.16)}$$

$$v(t+\partial t) = v(t+\frac{1}{2}\partial t) + \frac{1}{2}\partial t a(t+\partial t) \quad \text{.... (2.17)}$$

2.10. MD ensembles

A final consideration in setting up an MD simulation is the choice of statistical ensemble, such that a set of macroscopic parameters (N = number of atoms, V = volume of simulation cell, P = pressure, T = temperature) are fixed. The ensemble used in the following work is the canonical ensemble (NVT), where temperature fluctuates around a mean (preset) value via coupling of the system to a thermal bath which acts as a thermostat on the temperature. The thermostating process acts as a randomising force on the system, allowing more areas of configurational space to be sampled and ensuring ergodicity. Ergodicity is important as an MD run seeks to obtain time-averaged properties. We must make sure enough representative conformations are generated throughout the trajectory in order that the averages are meaningful. An NVE ensemble corresponds to an isolated (*i.e.* non-thermostated) system, and to generate good data a series of NVE simulations at different temperatures is needed. Maintaining a set temperature is difficult in this ensemble: it is necessary to start at a temperature approximately double the required, as the excess energy (temperature) equipartitions between both the kinetic and potential energies of the system. The energy partition is only exact for harmonic systems, however, and the unrealistic high starting temperatures can be damaging to the model.

2.11. Conclusions

The short introduction above shows the increasing variety of methods available to today's computational chemists. Continuing improvements, both in theory and also in computer power, have opened the door to understanding progressively more complex systems. Ten years ago ab initio quantum mechanical simulations were restricted to only a few atoms, now we can model cells containing hundreds of atoms on a realistic timescale and at an acceptable cost. The results presented in this thesis are testament to these advances, as results are shown for a 100-150 atom models for ca. 20 ps MD trajectories The project required considerable computing power, which was accessed through the UK's high performance computing facilities HPCx and HECToR. It is only through access to such state of the art software and hardware resources that the simulations presented in this thesis were rendered possible.

2.12 Bibliography

1. Young, D.C., *Computational Chemistry: a practical guide for applying techniques to real world problems.* 2001, New York: Wiley.

2. Jenson, F., *Introduction to Computational Chemistry*. 1999, New York: Wiley.

3. Martin, R.M., *Electronic Structure, basic theory and practical methods.* 2004: Cambridge University Press.

4. Payne, M.C., et al., *Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients.* Review of Modern Physics, 1992. **64**: p. 1045-1097

5. Gillan, M.J., *The virtual matter laboratory.* Contemporary Physics, 1997. **38**: p. 115-130.

6. Hohenberg, P. and W. Kohn, *Inhomogeneous Electron Gas.* Physical Review, 1964. **136**: p. B864-871

7. Kohn, W. and L.J. Sham, *Self-Consistent Equations Including Exchange and Correlation Effects.* Physical Review, 1965. **140**: p. A1133-1138

8. Perdew, J.P., K. Burke, and M. Ernzerhof, *Generalized Gradient Approximation Made Simple.* Physical Review Letters, 1996. **77**: p. 3865-3868

9. Perdew, J.P. and Y. Wang, *Pair-distribution function and its coupling-constant average for the spin-polarized electron gas.* Physical Review B, 1992. **46**: p. 12947-12954

10. Becke, A.D., *Density-functional exchange-energy approximation with correct asymptotic behavior.* Physical Review A, 1988. **38**: p. 3098-3100

11. Lee, C., W. Yang, and R.G. Parr, *Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density.* Physical Review B, 1988. **37**: p. 785-789

12. Ditchfield, R., W.J. Hehre, and J.A. Pople, *Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules.* The Journal of Chemical Physics, 1971. **54**: p. 724-728.

13. Hehre, W.J., R. Ditchfield, and J.A. Pople, *Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules.* The Journal of Chemical Physics, 1972. **56**: p. 2257-2261.

14. Vanderbilt, D., *Soft self-consistent pseudopotentials in a generalized eigenvalue formalism.* Physical Review B, 1990. **41**: p. 7892-7895

15. Hamann, D.R., M. Schlüter, and C. Chiang, *Norm-Conserving Pseudopotentials.* Physical Review Letters, 1979. **43**: p. 1494-1497

16. Schlegel, H.B., *Optimization of equilibrium geometries and transition structures.* Journal of Computational Chemistry, 1982. **3**: p. 214-218.

17. Shanno, D.F., *Conditioning of quasi-Newton methods for function minimization.* Mathematics of Computation, 1970. **24**: p. 647-656.

18. Goldfarb, D., *A family of variable metric methods derived by variational means.* Mathematics of Computation, 1970. **24**: p. 23-36.

19. Fletcher, R., *A new approach to variable metric algorithms.* Computer Journal,, 1970. **13**: p. 317-322.

20. Broyden, C.G., *The Convergence of a Class of Double-rank Minimization Algorithms.* Journal of the Institute for Mathematics and Applications, 1970. **6**: p. 222-231.

21. Car, R. and M. Parrinello, *Unified Approach for Molecular Dynamics and Density-Functional Theory.* Physical Review Letters, 1985. **55**: p. 2471-2474

Chapter 3:

The Poly-glycine Model

The construction of the basic poly-glycine scaffold with central water-wire and subsequent modeling of Proton Transport through the pore.

3.1 Introduction

The alpha-helix employed in this chapter is the simplest possible, consisting of a right-handed helix of a single amino acid, glycine (GLY, $NH_2CHRCOOH$, where R = H, the side-chain). There are no polar side-chain residues capable of interaction with the water-wire so this model will be both the benchmark, and basic scaffold, for all future simulations.

The key questions to be answered in the Chapter are:

- Is this model structurally stable and capable of supporting Grotthuss-type proton transport (PT)?
- How important are dispersion interactions to helical integrity?
- What is the effect of introducing an excess proton into a chain of water molecules?
- What effect does the density of water molecules have on the extent of PT and structure of the water-wire?
- What cationic species are supported in water-wires? And how does this vary with wires of differing length?
- What is the mechanism for PT in an alpha-helical environment and how does it compare to bulk water?
- What effect does the choice of functional have on water-wire structure and the extent of PT?

3.2 Simulation Methods

Model: A polypeptide chain of glycine was twisted into a right handed alpha helix (with standard Phi and Psi angles of -57.0° and -47.0°, respectively) using the ARGUSLAB[1] polypeptide builder. The simulation was run under full periodic boundary conditions (PBCs) by placing one full repeat of the helix (*i.e.* four helical turns, comprising fifteen amino acid residues) at the corner of an orthorhombic cell to lie along cell vector *b*, corresponding to 26.36 Å (see Figure 3.1). Thus cell vector *a* (= *c*,

initially set to 8.0 Å) defines the diameter of the inter-helical pore, into which the water environment (comprising in turn eight, ten and twelve water molecules, plus an extra proton) is placed. In this way the one helix-model is replicated to four, and in turn this packing arrangement is repeated to infinity along the *a,b* and *c* vectors.

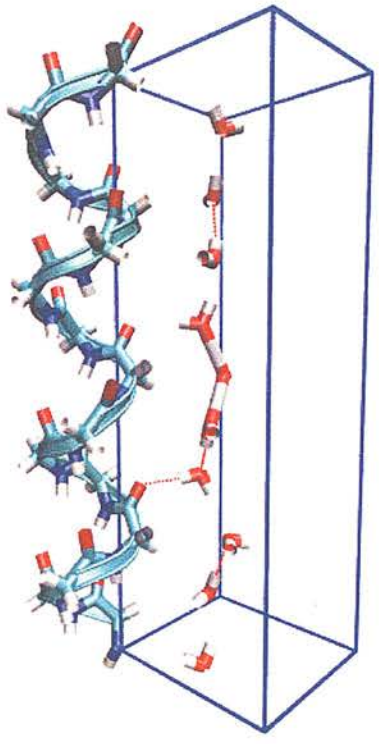


Figure 3.1 The ten molecule water-wire simulation model.

Simulations: All simulations were run using the CP2K[2, 3] molecular dynamics simulation package with the GGA Becke-Lee-Yang-Parr (BLYP) [4, 5] functional, coupled to a dual localised (Gaussian) and plane-wave basis set description. The localised basis set was of double-zeta quality and optimised for use against the Goedecker-Teter-Hutter set of pseudopotentials [6] coupled to the BLYP functional. A series of single-point energy calculations determined the optimum energy cut-off (300 Ry), and the subsequent geometry optimisation was performed in two steps using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) [7-10] method. In the first instance atom positions only were allowed to relax; this was then followed by a series of single-point energy calculations to obtain the optimised cells vectors, during which the *b* vector

shortened slightly (to 26.26 Å) while the *a* and *c* vectors remained unchanged. The water-wire model was then constructed by adding an extra hydrogen atom to one of the water molecules to create an oxonium ion, $[H_3O]^+$. The eight and twelve water-wire models were then built from this baseline model, by removing or adding two further water molecules, respectively. All three models were then subjected to atom minimisation, which then formed the starting-points for NVT *ab initio* molecular dynamics equilibration simulations. Dispersion, or van der Waals, interactions, which are inherently missing from DFT calculations, were accounted for by the addition of a pair potential.[11] The MD simulations (maintained at 365K by a chain of Nose-Hoover thermostats) were run using the same basis set as described above for *ca.* 25 ps, advancing in time increment steps of 0.55 fs. After these equilibration runs, the production runs (of at least 25 ps) were then obtained under NVE conditions for each model. Visualisation of the models and the MD trajectories was performed using the VMD package. [12]

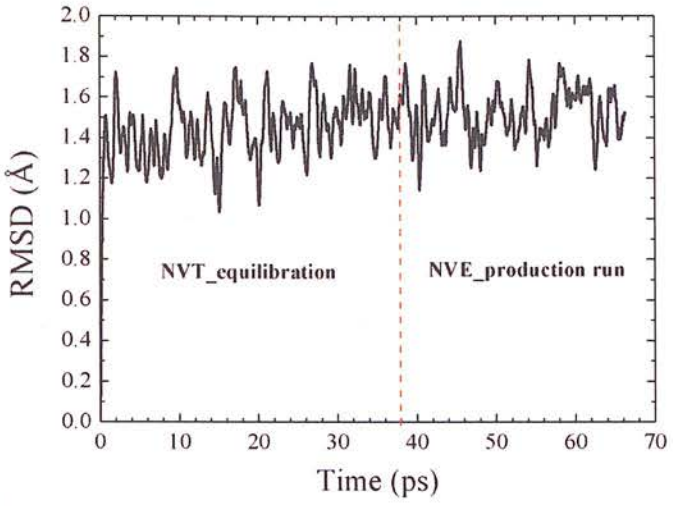
3.3. Results and Discussion

3.3.1. Structural stability of the model

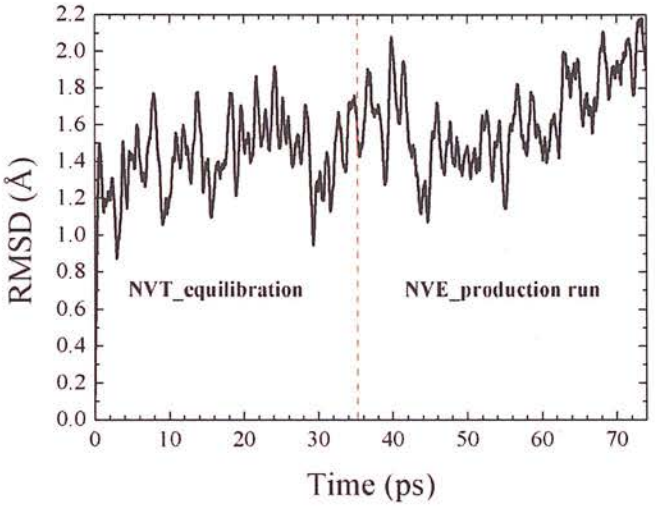
In this study the model used is a single alpha-helix and water environment under PBCs with no atomic positional restraints, and thus the stability of the model must be checked. The integrity of the PBC along the *b* vector was tested by verifying that the time-average C-N bond (which forms the continuous helix by bridging the cell boundary) is consistent with the time-averaged C-N bonds present in the rest of the helix. We compared the boundary value with three other C-N bonds presented in other parts of the helix, and all were found to be identical to within one standard deviation. This establishes that the boundary conditions remained intact throughout the simulation, and the channel can be modelled as continuous.

Figure 3.2 shows the root mean square deviation (RMSD) of the helix backbone over the course of two trajectories for our baseline model (comprising ten water molecules and an extra proton), relative to the initial (t = 0, *i.e.* optimized) structure. The plots show the importance of including the dispersion correction to prevent the

helix from unravelling. With the correction in place [Figure 3.2(a)] the atom positions are shown to fluctuate noticeably, with an average displacement of 1.4(2) Å (and a maximum of 1.9 Å). Although there is no recognised standard or upper limit for the RMSD, previous studies are comparable to this one, reporting RMSDs in the range 1-1.5 Å. [13] Crucially, most of the deviation occurs within the first picosecond of the simulation, after which the plot levels off which indicates that the helix does not unfold during the course of the simulation. In the absence of the dispersion correction [Figure 3.2(b)] a slightly higher average RMSD was obtained [1.5(3) Å with a maximum value of 2.2 Å]. However, this plot does not level off after the 25 ps equilibrium period, indicating likely breakdown of the helix in this case. Thus we conclude that the dispersion interactions play a crucial role in stabilizing the alpha-helical structure, which echoes recent work on the DNA-duplex. [14]



(a)



(b)

Figure 3.2 The relative mean square displacement (RMSD) of the helix backbone atoms for the ten water-wire model with (a) dispersion corrections included and (b) no dispersion correction.

The channel radius as a function of the *b* axis was further investigated using the HOLE program, [15] which maps the internal cavity of the channel based on the maximum radii of a series of hard spheres. The resulting representation is given in Figure 3.3, were the radii of the hard spheres was observed to fluctuate around 2.1(2) Å

for most of the channel. A minimum radius of 1.15 Å [16] is required to accommodate a water molecule, thus indicating that all parts of this channel are accessible to the water-wire.

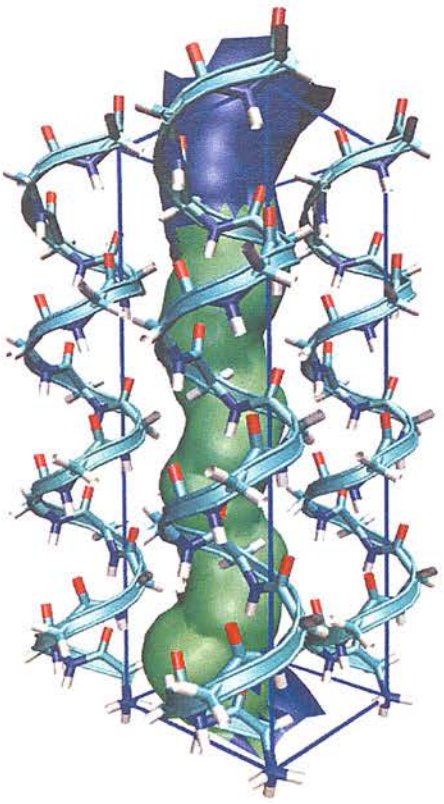


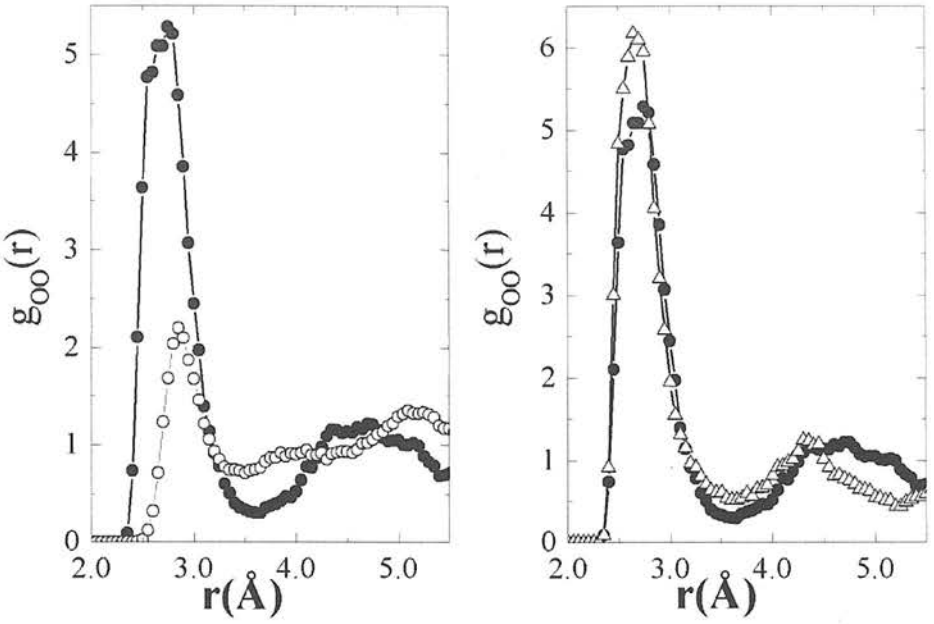
Figure 3.3 The simulation model showing the internal cavity of the simulation cell, into which eight, ten or twelve water molecules plus an extra proton are placed

3.3.2. Behaviour of water wire vs. water chain

The effect of introducing a proton to the chain of ten water molecules was investigated by a direct comparison of the oxygen-oxygen radial distribution functions (RDFs), $g_{OO}(r)$, obtained for the water-chain and water-wire systems. These plots, shown in Figure 3.4(a), provide information on the nature of the hydrogen bonding between the water molecules. Our results echo those of Voth *et. al.* [13] obtained from studying PT in cylindrical, smooth-walled hydrophobic channels. Considering first the O...O separations found for the water-wire, the dominant peak lies at 2.75 Å with

density starting at 2.4 Å. Gas phase calculations have shown an average donor-acceptor separation of around 2.4 Å for the Zundel complex, $[H_5O_2]^+$, and 2.57 Å for the Eigen complex, (which is a central $[H_3O]^+$ ion solvated by three water molecules). On this basis we assign the shortest separations observed to the formation of the Zundel complex. Although we do see density at *ca.* 2.6 Å we cannot attribute this to a standard Eigen complex as the channel width is too narrow to accommodate this type of solvation structure. We do, however, observe many instances of an Eigen-type structure with O...O separations of around this value, where the hydrogen bond of the third water in the bulk is replaced by a contact to a carbonyl oxygen of the helix backbone to form a 'tethered' Eigen complex. Studies of excess protons in bulk water environments have suggested that of the two species the Eigen complex is the more stable. [17] Thus it would appear that the formation of the tethered Eigen complex in our model channel represents an attempt to mimic this stable species.

The O...O separations observed on the RDF from the simulation without the excess proton show a general shift to longer separations (2.85 Å) between the water oxygen atoms, which is consistent with that reported on a g(r) plot of bulk water. Comparison of the $g_{OO}(r)$ plots suggests that the effect of the excess proton extends beyond those water molecules immediately surrounding it and leads to a structuring of the water molecules in the next solvation shell, which is in agreement with previous findings. [18, 19] This feature is absent on the water-chain plot, thus suggesting that in the absence of the extra proton the water molecules are considerably less organised and adopt a wider distribution, presumably due to the increased mobility of the less-coordinated channel water molecules.



(a) (b)

Figure 3.4 Radial distribution functions $g_{OO}(r)$ obtained from (a) the ten water molecule model (OO) (open circles) and in the presence of an extra proton (O*O) (filled circles) using the BLYP functional and (b) $g_{OO}(r)$ in the presence of an excess proton (O*O) for the BLYP (filled circles) and PBE functionals (open triangles).

3.3.3. Effect of changing the density of water molecules in the channel

The simplified model employed in this work cannot support a water reservoir at either end of the inter-helical pore. The simulation must therefore start with a pre-determined number of water molecules already present in the pore, and so the effect of varying the density of water must be carefully checked. A comparison of the RMSDs for the eight, ten and twelve water-wire models appears to suggest that altering the density of water molecules in the central cavity does influence the behaviour of the helical scaffold. As the number of water molecules in the cavity is increased the average RMSD increases from 1.4 to 1.7 Å. This is perhaps not too surprising: the helix is over 26 Å long, and interactions with water molecules could easily activate low-energy transverse oscillations that are picked up in the RMSD plots.

Time-averaged plots obtained over 25 ps NVE trajectories for the eight, ten and twelve water-wire models are given in Figure 3.5. Concentrating first on the ten-membered water wire [Figure 3.5(b)] it is apparent that for the majority of the simulation at least five of the water molecules hydrogen-bond to carbonyl groups on the alpha-helices. The excess proton travels over five (though predominantly four) of the ten water molecules, covering a distance of approximately 7.5 Å, or a quarter of the channel length, without any significant diffusion of the water molecules themselves. The pattern of hydrogen bonding within the wire and around the excess proton is seen to radiate from a central solvated oxonium ion, with the central water molecules participating in three hydrogen bonds (two with neighboring water molecules with a third contribution from a backbone carbonyl oxygen) and the outer in two (due to the loss of one hydrogen bond with a water molecule). The ordering of water molecules within the water-wire was further investigated using the average atomic distributions of water O- and H-atoms along the poly-GLY channel axis, as shown in Figure 3.6. The plot has two distinct regions, with those water molecules involved in PT easily distinguished from those which are not. Strong hydrogen bonding within the wire is indicated by hydrogen atom peaks that are placed between successive oxygen atoms to form a regular donor-acceptor pattern [particularly so for oxygen atoms 6-9, see Figure 3.6(b)]. These four peaks show the preferred position of the excess proton, which is found in a conformational spread over four oxygen atoms. The donor-acceptor pattern radiates out from this central area, holding these four water molecules in a tight configuration, as indicated by the sharpness of the oxygen peaks. There is no diffusion here, but simply fluctuation around a set position. Five of the remaining water molecules (2-5,and 10) are seen to form a water chain separate from the water-wire, with the wider peaks indicating these water molecules have an increased mobility compared to those in the water-wire.

A reduction in the number of water molecules in the wire has a very noticeable effect on its structure and dynamics [see Figures 3.5(a) and 3.6(a)]. The three central peaks of the O- and H-atom distributions again show the preferred position of the excess proton, which in this case is delocalised over two water molecules and a carbonyl oxygen on the alpha-helix to form an Eigen-type cation. Again the donor-acceptor pattern radiates from the centre, but in this case the order extends only to the

central five water molecules. The rest are relatively mobile. The removal of two water molecules has therefore resulted in a further break in the water-wire, rendering a substantial volume of the channel dry. Proton transport cannot bridge this gap, which limits the distance the proton can travel to only 3.5Å.

Adding two water molecules to the baseline ten water-wire model also alters the behavior of the system significantly: the proton is now delocalised over six to seven water molecules, and travels a distance of 10.5 Å [See Figures 3.5(c) and 3.6(c)]. The time-averaged structure, however, shows reduced hydrogen bonding between the water-wire and the helix carbonyl groups with only three water molecules shown to participate in bonding. The O- and H-atom distributions show five or six water molecules held strongly around the excess proton, while those beyond this sphere of influence are considerably more mobile, particularly water molecules 3-5. The time-averaged structure [Figure 3.5(c)] shows a small break in the water-wire between molecules 1 and 2 and 2 and 3. The O- and H-atom distributions, however, do not show this same feature, but do indicate a greatly increased mobility, which when averaged out may result in this artifact.

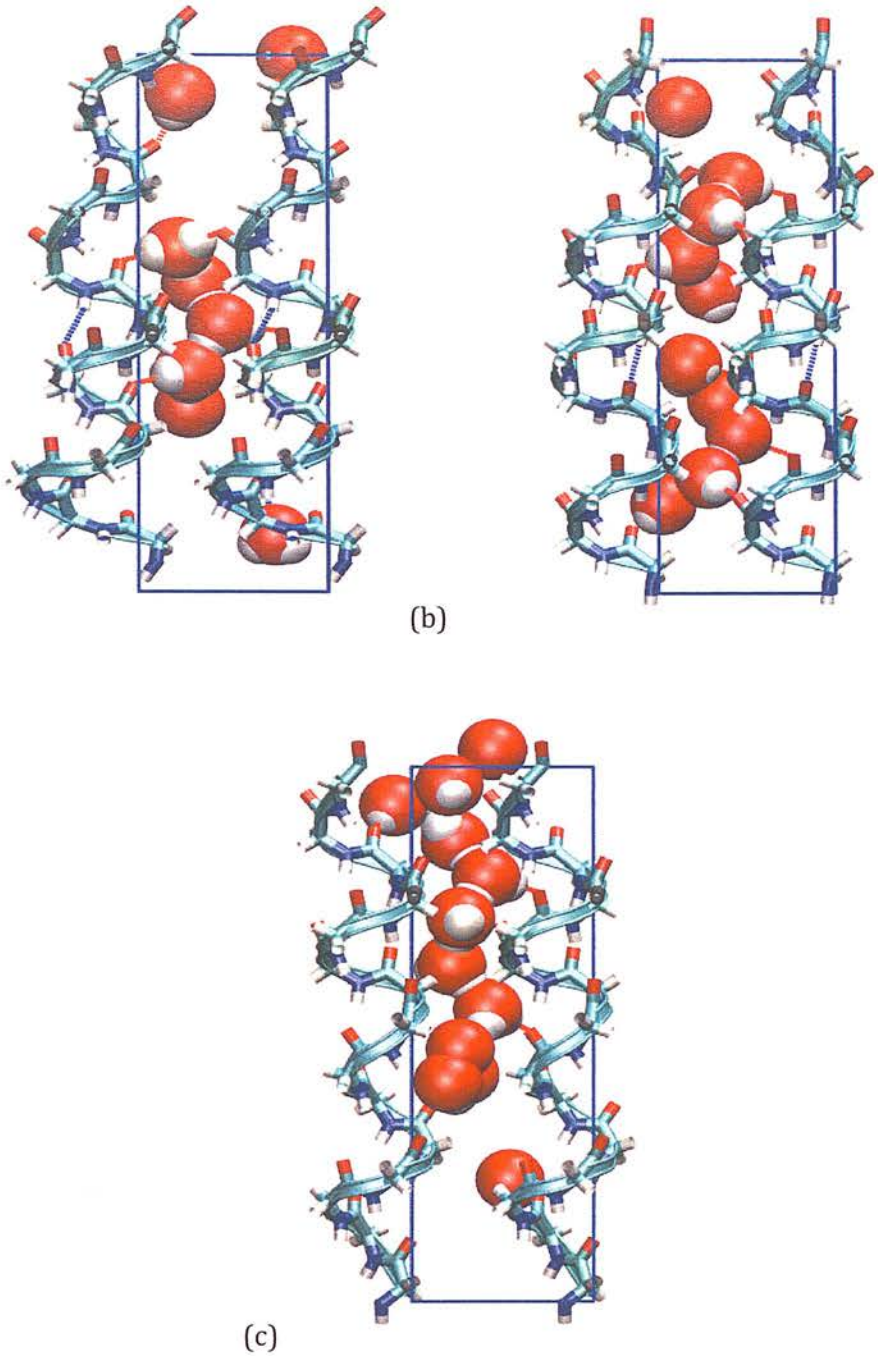


Figure 3.5 Time-averaged plots of the (a) eight-, (b) ten-, and (c) twelve-water-wire models. Note the water molecule numbering scheme used in the text is defined with respect to the bottom of each of these simulation cells.

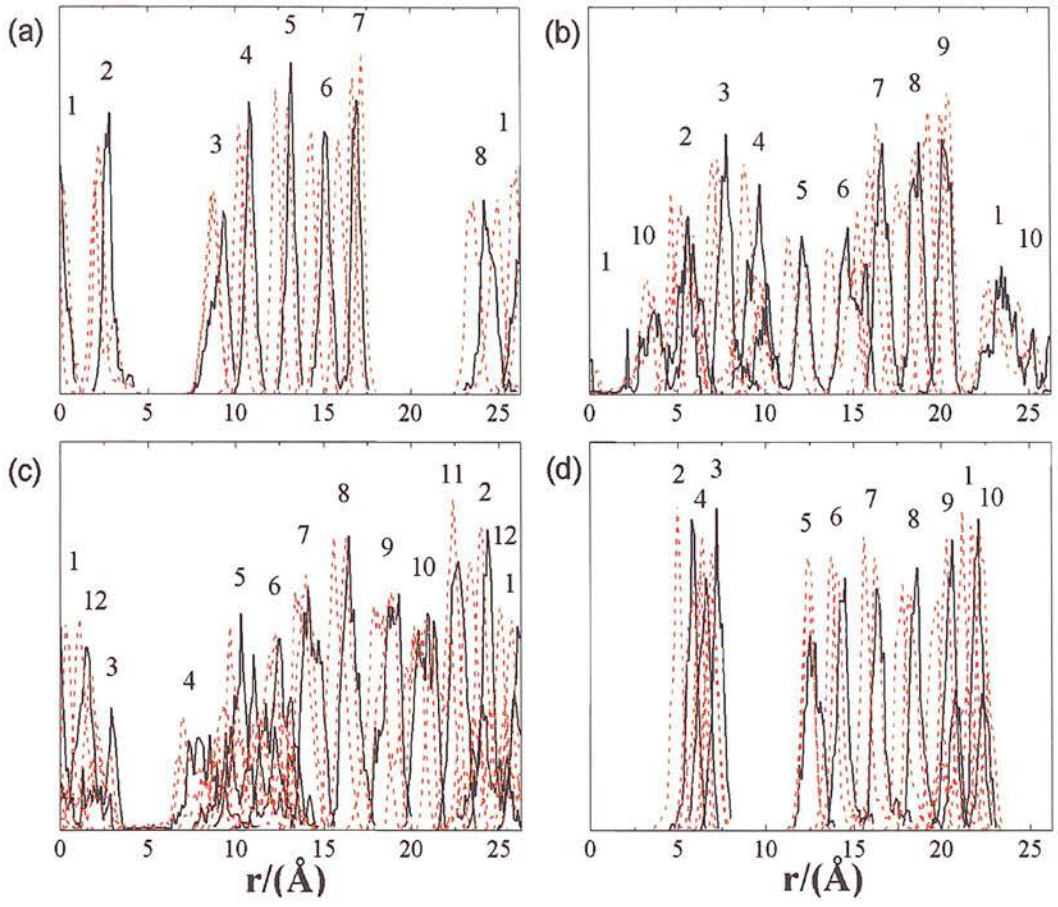


Figure 3.6 Average atomic distributions of water oxygen (black solid line) and hydrogen (red dashed line) atoms modelled along the channel axis for the (a) eight-, (b) ten-, and (c) twelve- water-wire models. Also shown is the (d) ten-water wire model system using the PBE functional.

3.3.4. Calculating the lifetimes of cationic species formed during the MD simulation

The average lifetimes of the different cationic states (*i.e.* $[H_{2n+1}O_n]^+$, where n = 2-10) that were created during the MD trajectories can be obtained by defining a time autocorrelation function for bond existence as

$$C_e(t) = \frac{1}{N}\sum_{i=1}^{N}\langle \alpha(0) \times \alpha(t) \rangle \qquad \text{.... (3.1)}$$

where $\alpha(t)$ is defined to be 1 if the interaction condition between a pair of atoms is met at time t, and 0 if it is not. A hydrogen bond was defined by considering interactions between oxygen atoms that fell within a cut-off distance of $R_{OO} \leq 3.5$ Å, which was chosen to encapsulate the first peak on the $g_{OO}(r)$ plot shown in Figure 3.4(a), along with an H-O...O angle constraint of 30°. The first step in the process is to identify the oxygen atom which has the three closest hydrogen atoms, *i.e.* the oxonium ion (denoted O*). From there, to track a Zundel cation $[H_5O_2]^+$ the O...O distance from O* to a neighbouring water molecule must be less than or equal to 2.75 Å, and these in turn must be separated from all other water molecules by a distance greater than 3.5 Å. Extensions to this methodology then allow for the higher hydrated cation species (*i.e.* the proton delocalised over three, four or five water molecules) to be similarly investigated. Integration of the resulting time autocorrelation functions then yield the corresponding cation lifetimes. In addition the production run MD trajectories were analysed to count the number of instances that the geometric criteria for the existence of the individual cationic states were fulfilled. The results, along with the lifetimes, are given in Table 3.1. From this we see that there is a wide range in cationic state lifetimes, from tens of femtoseconds, to several picoseconds. We also observe that there is a rough correlation between the lifetimes and the prevalences for each state, *i.e.* in general the more times the state is formed, the longer is its lifetime. At no point was an isolated Zundel (n = 2) cation observed, however the $n > 3$ states may contain the Zundel structure solvated by $n - 2$ water molecules. All results are consistent with the time-averaged plots in Figure 3.5 and the average water distributions in Figure 3.6.

The baseline ten water-wire model supports the formation of water-wires from n = 3-9, but shows a clear preference for the four-membered water-wire state, (*i.e.* $[H_9O_4]^+$). This state has a very long lifetime of 2.6 ps, reinforcing this as the most stable cationic state for the ten water-wire model. The reason for this lies with the fact that the water-wire in this model broke half-way through the simulation, after which time the simulation essentially settled into the $[H_9O_4]^+$ state. The observation of states n = 7-9 with lifetimes of several hundreds of femtoseconds relate to the first half of the

simulation before the wire broke; intermediate states of n = 5-6 were also observed but with very short lifetimes that correspond to the timescale of the O-H bond stretching vibration.

Table 3.1 Prevalences (expressed as a % of the total simulation time) and lifetimes (in fs) of the cationic states $[H_{2n+1}O_n]^+$ observed in the eight-, ten- and twelve water-wire models.

	8 water-wire		**10 water-wire**		**12 water-wire**	
n	*Prevalence (%)*	*Lifetime (fs)*	*Prevalence (%)*	*Lifetime (fs)*	*Prevalence (%)*	*Lifetime (fs)*
3	2.4	32	0.5	9	0.01	4
4	1.4	10	43.7	2569	0.7	64
5	96.2	19	4.7	53	3.6	151
6	-	-	2.2	22	31.3	964
7	-	-	12.3	942	17.4	489
8	-	-	18.6	210	6.2	196
9	-	-	18	517	13.2	400
10	-	-	0.1	10.2	27.9	774

The eight water-wire model distribution peaks at n = 5, the creation of which was observed for over 95% of the simulation. Wires above five water molecules in length are not observed in the trajectory, as the loss of two water molecules creates two breaks in the overall water-wire and the excess proton is confined to the central region of the channel [Figure 3.6(a)]. The lifetimes for this model indicate that the system is extremely fluxional, with all species decaying on the femtosecond timescale. Finally, the twelve water model shows a preference for the formation of longer water-wires, peaking at n = 6 and n = 10.

3.3.5. The mechanism for PT in a model alpha-helical channel

Recent work on bulk water simulations[19, 20] has made use of $g_{OH}(r)$ RDFs centred on the oxonium oxygen (O*), the oxygen which will accept the proton (O_{next}) and its other nearest neighbours ($O_{nearest}$) averaged over the whole simulation to try and elucidate a mechanism for the PT process. A successful PT event is defined as one

where the next transfer event does not see the proton simply reverse its direction. These reversals are unsuccessful and are termed 'rattling' events. For direct comparison with this work, we present our data for the ten water-wire model in this format. Figure 3.7 shows a series of time-dependent $g_{OH}(r)$ RDFs centred on O_{next} and $O_{nearest}$ at four time intervals leading up to a PT step; (a) well before the event and (b)-(d) the time directly preceding it. An average is taken over all PT steps, with the rattling events excluded.

For ease of discussion we consider first the $O_{nearest}$ g(r) plot (dashed black line, Figure 3.7). The dominant, sharp, peak is centred at 1.1 Å, which can be assigned to the hydrogen atoms directly bound to the oxygen atoms to form a water molecule. A second, broader peak is centred at around 1.7 Å, which corresponds to an O...H hydrogen bond interaction, and a third broader peak at around 3.3 Å which is an O...H distance over a wider coordination sphere. This plot remains essentially constant throughout the four time intervals, indicating that the $H_2O_{nearest}$ molecule does not play a role in the PT mechanism. The O_{next} g(r) plot (solid black line, Figure 3.7), does however show important changes over the time intervals shown. The three peaks are again present: the dominant peak at 1.1 Å, the second peak now appears as a shoulder at *c.a.* 1.4 Å, and the third peak is now sharper and centred around a slightly shorter distance of *c.a.* 3.1 Å. The fact that the second peak appears at a shorter distance makes perfect sense: the $[H_3O]^+$ and H_2O_{next} molecules are closer together, hence the heightened incidences of PT rattling events. During the time period leading up to the PT event this peak shifts to the left (from 1.4 Å to 1.3 Å) demonstrating the formation of the short, strong hydrogen bond indicative of a Zundel cation, $[H_2O_5]^+$. This peak shift agrees with previous findings, [20] but we acknowledge that our data are less clear, as our second peak appears as a shoulder rather than a distinct second peak.

In our channel model, all successful PT steps are preceded by a large number of unsuccessful or 'rattling' PT events as described above. This is shown in Figure 3.8(a), where the identity of the oxonium ion *versus* time for the whole of the MD production run is displayed in the same manner as Markovitch *et. al.* for bulk water. [19] The rattling events are evident from the large number of incidences where the O* identity repeatedly flips with a neighbour (O_{next}). A plot leading up to one particular PT event that highlights the representative time intervals presented in Figure 3.7 is shown in Figure 3.8(b), where the constant fluctuations prior to the PT event can be clearly seen.

The large number of rattling events can be traced back to the fact that the direction of PT is limited in the channel environment and the positions of the water molecules are inevitably constrained by the presence of the helix. To continue the comparison with bulk water, we also track the identity of O_{next} [shown in Figure 3.8(c)] over the same time interval shown in Figure 3.8(b). In bulk water the excess proton resides on one particular water molecule for sustained periods of time, which allows for detailed analysis of the corresponding behaviour for $O_{next}/O_{nearest}$. During these stable periods it is possible to see fluctuations in O_{next} identity, such that the three water molecules in the first solvation shell compete to get closest to O*. This has been termed the 'special pair dance'. [19] In our channel environment the heightened incidences of rattling mean we do not observe the same long periods of time where the identity of O* remains fixed. During the short time periods [Figure 3.8(b)] where O* does not change there is some evidence of O_{next} flipping [Figure 3.8(c)]. However, the statistical certainty is lacking to state conclusively whether the 'special pair dance' phenomenon found in bulk water also exists in the channel simulations.

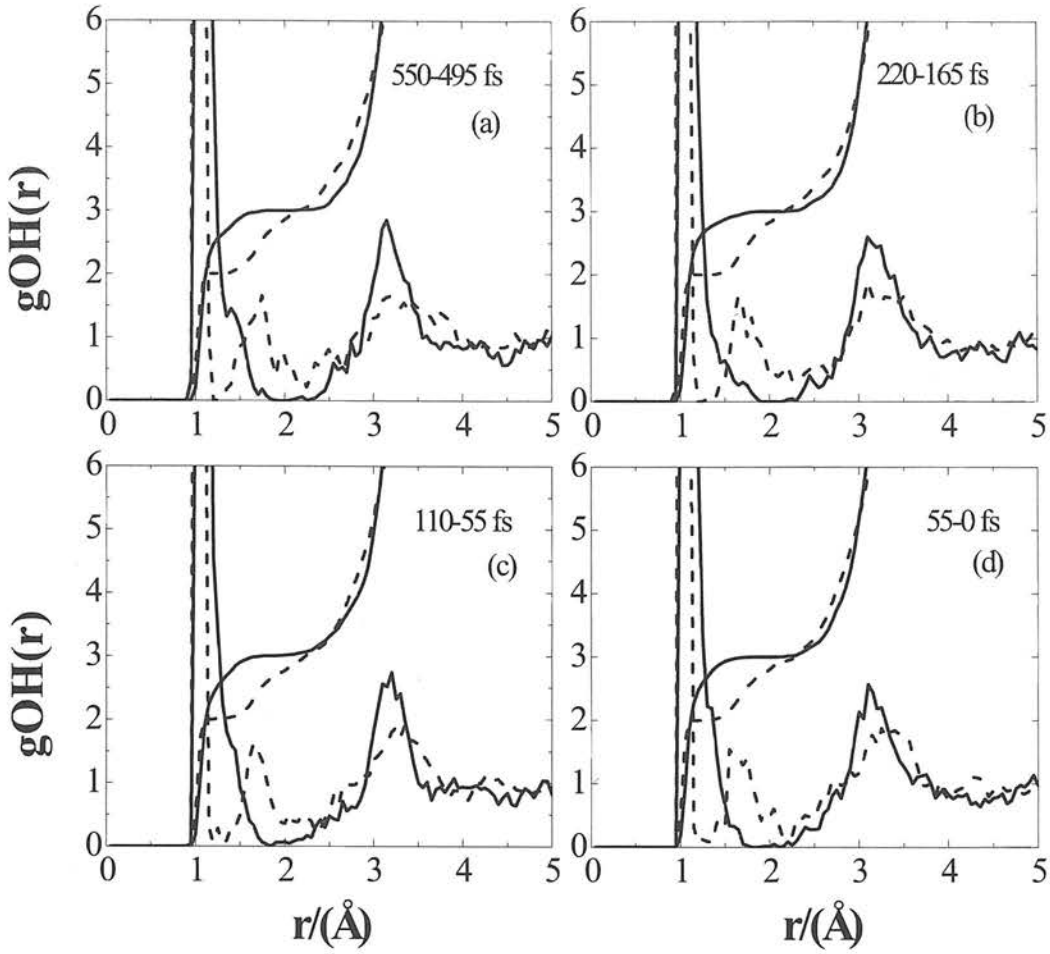


Figure 3.7 Time dependent RDFs and integrated coordination number plots of O_{next} (solid black line) and $O_{nearest}$ (dashed black line) averaged over 55 fs intervals before [(a) 550-495 fs] and approaching a PT event [(b) 220-165 fs, (c) 110-55 fs and (d) 55-0 fs]. PT occurs at time 0 in panel (d). Also included in the graphs above are integration curves which indicate the number of hydrogen atoms at each O-H separation.

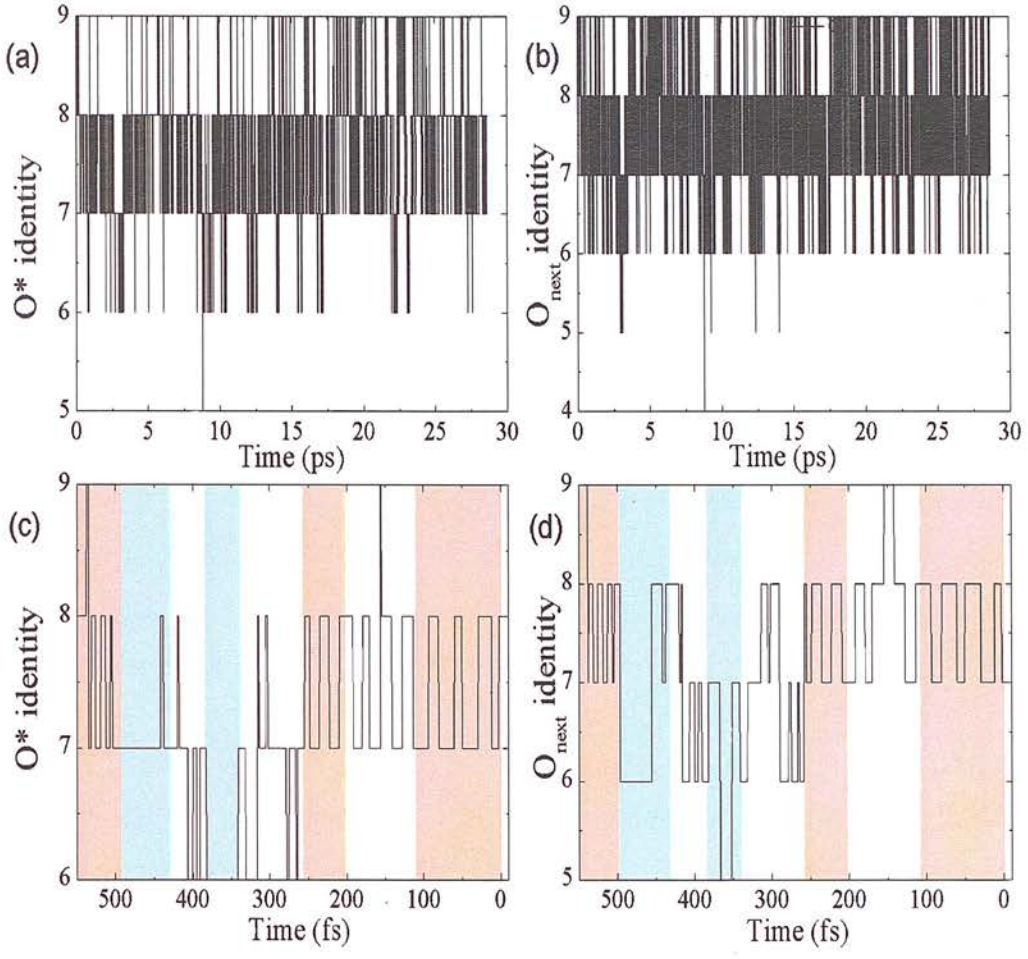


Figure 3.8 Plot of oxonium identity *vs* time for (a) the whole MD production run, (c) the time interval leading up to a PT event, and (b) the identity of O_{next} over the entire production run and (d) O_{next} during the interval leading up to a PT event. The pink shaded panels correspond to the time intervals given in Figure 3.8(a)-(d), the cyan to regions where the identity of O* remains fixed for short periods of time.

One further significant deviation from the bulk water mechanism of PT can be identified. In bulk water simulations a key step to successful PT is the Eigen-Zundel-Eigen isomerisations (E-Z-E) [21, 22] which indicates the importance of hydrogen bond cleavage to successful PT. The previous work [20] validated this E-Z-E theory as analysis of the oxygen atom coordination number clearly showed the breaking and formation of hydrogen bonds along the PT trajectory: O_{next} looses a hydrogen bond (the coordination drops from four to three), and O* gains a hydrogen bond (the coordination rises from three to four). A channel environment cannot support the same characteristics as it imposes natural limits on the number of hydrogen bonds a water

molecule can form. In our simulation a maximum value of three is observed (from the integration plot in Figure 3.7), *i.e.* two hydrogen atoms are directly coordinated to the oxygen atom and a third is accepted from a neighbouring water molecule. Thus we can conclude that the water molecules in the channel are coordinatively unsaturated. In the case of the close hydrogen bonded network involved in PT this coordination number will not change, *i.e.* unlike bulk water the breaking of a hydrogen bond is not a precursor to a successful transfer event. This study therefore highlights an important difference in the PT mechanism for a channel environment compared to bulk water.

Finally, as our model is more chemically relevant than the work that has gone before us, [13, 23-27] we should be able to report more on the properties specific to an alpha-helical channel environment. We have already observed some evidence that as the density of water in the channel increases the amount of helical movement also increases. Looking more closely at the time-averaged pictures in Figure 3.5 also suggests that the active PT regions of the water-wire correspond with those areas of the helix that are most tightly coiled. Or conversely, water diffusion behaviour is observed over those regions where the helix has undergone the most visible change. This is particularly noticeable in the top section of the eight water-wire model [Figure 3.5(a)] and the bottom section of the twelve water-wire model [Figure 3.5(c)]. This would suggest that when there is a clearly defined helical groove the water molecules will adopt favourable positions that are capable of displaying longer-range cooperative behaviour and thus can support PT.

3.3.6. The effect of changing the DFT functional on the behaviour of the water wire

It is well known that current GGA functionals do not accurately describe the behaviour and properties of water molecules in the bulk, with *e.g.* diffusion rates and hydrogen bond donor-acceptor distances not correctly represented during the MD simulation, [28] although recent evidence would suggest that the use of a complete basis set greatly improves the properties mentioned above. [29] There is however little documentation on the effects of different functionals on the properties of water molecules within a water-wire. To address this point, further MD simulations on the baseline ten water molecule model with and without an extra proton were carried out

using the PBE functional as expressed in the CP2K package, with all other conditions consistent with the BLYP baseline model reported above. Analysis of the trajectories revealed that there are significant differences in the behaviour of the wire with the choice of functional.

Comparison of the $g_{OO}(r)$ plots of the ten water-wire baseline model MD simulations using BLYP and PBE functionals show a distinct difference in the structuring and behavior of the water-wire [compare Figure 3.4(a) with 3.4(b)]. The PBE plot shows a shift to shorter OO separations, with the distribution peaking at 2.65 Å compared to the 2.75 Å maximum for the BLYP model, which is itself in good agreement with experiment. [30] The secondary solvation shell is much more distinct in the PBE wire-model with a narrower, more defined peak at 4.3 Å as opposed to the larger spread observed for the BLYP model.

The O- and H-atom distribution plots for the two functionals again illustrate the effects of varying the functional on the formation and dynamics of the proton wire [compare Figure 3.6(b) and 3.6(d)]. The PBE model demonstrates an increased structuring of the water molecules, with all but one water molecule held in a tight conformation. The influence of the excess proton is increased to six water molecules, but those molecules that have broken off from the main water-wire are themselves held in an extremely tight hydrogen-bonding pattern, as shown by the narrow, sharp distribution.

We also note that from a structural chemistry perspective the orientation of the water molecules within the wire in the PBE model are in more optimal positions for PT than observed in the BLYP simulation, and yet the overall degree of PT observed with this functional is suppressed. While the proton does travel over six of the water molecules, around 96% of its time is spent on O(9) and O(8). There are therefore significantly fewer successful PT events and greatly increased unsuccessful rattling events. From these results we therefore conclude that the BLYP functional appears to provide a more dynamical description of the water-wire than PBE. Previous work tested the accuracy of common density functional methods using six- and eight-water clusters.[31] Using MP2 calculations as a benchmark they found that while the PBE functional provided a closer estimate of the geometry, the BLYP functional calculated the energy to a greater level of accuracy. Overall the hybrid functional B3LYP was

suggested as the best DFT functional for this purpose. However, from an *ab initio* molecular dynamics perspective we note that any higher level functional that makes use of the Fock matrix for exact exchange is likely to remain intractable when coupled to a plane-wave basis set for some time to come.

3.4. Conclusions

The answers to the questions raised in the introduction of this Chapter have now been addressed. The simple model we proposed for a trans-membrane protein channel that supports PT is structurally stable to ab initio geometry optimization and molecular dynamics simulation. Including dispersion interactions are crucial to helix stability. The model supports proton transport, with the maximum distance travelled for an excess proton along a chain of water molecules in the absence of any external driving force is approximately 10.5 Å, which corresponds to around 1.5 helical turns. The addition of an excess proton to a water chain was observed to set up long-range cooperativity between water molecules, with evidence of the formation of second solvation shells observed in g(r) plots. This is consistent with previous findings. The effect of altering the number of water molecules within the channel was also investigated. Some evidence was observed that the water molecules exerted some influence on the behaviour of the helix, with the RMSD for the model increasing slightly with water density. It was also observed that the regions of the channel that supported PT, rather than water diffusion, were those that were more tightly coiled. Within the wire a range of cationic species were observed, with lifetimes recorded over a wide range (from 10 fs to 2.6 ps).

The mechanism for PT in this alpha-helix channel model showed significant differences to that observed for bulk water. The Eigen-Zundel-Eigen isomerisations necessary for the propagation of an excess proton through bulk water are not possible in the channel environment. We observe that the number of coordinated hydrogen

atoms around an oxygen atom remains constant at three, *i.e.* it is not necessary to break a hydrogen bond for PT to occur as has been reported in bulk water simulations. Instead, there is a heightened incidence of proton 'rattling' events throughout the simulation, where the proton constantly flips between two neighbouring molecules.

Consequently, the amount of time the excess proton stays coordinated to any one particular water molecule is reduced, and the statistical certainty to state whether or not the bulk-water 'special pair dance' phenomenon occurs in the channel environment could not be obtained. In common with the bulk water mechanism, however, in the lead up to PT, we do observe a shortening of the O* to O_{next} hydrogen bond distance indicative of the formation of a Zundel complex.

Finally, we comment on the use of different computational models for an alpha-helix water-wire system. Our results suggest that the PBE functional over-structures the water-wire and suppresses PT whilst the DFT functional BLYP provides a more dynamical water-wire system. Our results also demonstrate the importance of including dispersion interactions in DFT simulations containing alpha-helical structures.

3.5. Bibliography

1. Thompson, M.A., *ArgusLab 4.0.1*, Planaria Software LLC: Seattle, WA.

2. The CP2K Developers' Group. 2008, http://cp2k.berlios.de

3. VandeVondele, J., et al., *Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach.* Computer Physics Communications, 2005. **167**(2): p. 103-128.

4. Becke, A.D., *Density-functional exchange-energy approximation with correct asymptotic behavior.* Physical Review A, 1988. **38**(6): p. 3098-3100

5. Lee, C., W. Yang, and R.G. Parr, *Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density.* Physical Review B, 1988. **37**: p. 785-789.

6. Hartwigsen, C., S. Goedecker, and J. Hutter, *Relativistic separable dual-space Gaussian pseudopotentials from H to Rn.* Physical Review B (Condensed Matter and Materials Physics), 1998. **58**(7): p. 3641-3662.

7. Broyden, C.G., *The Convergence of a Class of Double-rank Minimization Algorithms.* Journal of the Institute for Mathematics and Applications, 1970. **6**: p. 222-231.

8. Fletcher, R., *A new approach to variable metric algorithms.* Computer Journal,, 1970. **13**: p. 317-322.

9. Goldfarb, D., *A family of variable metric methods derived by variational means.* Mathematics of Computation, 1970. **24**: p. 23-36.

10. Shanno, D.F., *Conditioning of quasi-Newton methods for function minimization.* Mathematics of Computation, 1970. **24**: p. 647-656.

11. Grimme, S., *Semiempirical GGA-type density functional constructed with a long-range dispersion correction.* Journal of Computational Chemistry, 2006. **27**(15): p. 1787-1799.

12. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics.* Journal of Molecular Graphics, 1996. **14**: p. 33-38.

13. Brewer, M.L., U.W. Schmitt, and G.A. Voth, *The Formation and Dynamics of Proton Wires in Channel Environments.* Biophysical Journal., 2001. **80**: p. 1691-1702.

14. Rezac, J. and P. Hobza, *On the Nature of DNA-Duplex Stability.* Chemistry - A European Journal, 2007. **13**(10): p. 2983-2989.

15. Smart, O.S., et al., *HOLE: A program for the analysis of the pore dimensions of ion channel structural models.* Journal of Molecular Graphics, 1996. **14**: p. 354-360.

16. Smart, O.S., J.M. Goodfellow, and B.A. Wallace, *The pore dimensions of gramicidin A.* Biophysical Journal, 1993. **65**: p. 2455-2460.

17. Schmitt, U.W. and G.A. Voth, *The computer simulation of proton transport in water.* The Journal of Chemical Physics, 1999. **111**: p. 9361-9381.

18. Markovitch, O. and N. Agmon, *Structure and Energetics of the Hydronium Hydration Shells.* The Journal of Physical Chemistry A, 2007. **111**(12): p. 2253-2256.

19. Markovitch, O., et al., *Special Pair Dance and Partner Selection: Elementary Steps in Proton Transport in Liquid Water.* The Journal of Physical Chemistry B, 2008. **112**(31): p. 9456-9466.

20. Berkelbach, T.C., H.-S. Lee, and M.E. Tuckerman, *Concerted Hydrogen-Bond Dynamics in the Transport Mechanism of the Hydrated Proton: A First-Principles Molecular Dynamics Study.* Physical Review Letters, 2009. **103**(23): 238302.

21. Agmon, N., *The Grotthuss mechanism.* Chemical Physics Letters, 1995. **244**: p. 456-462.

22. Lapid, H., et al., *A bond-order analysis of the mechanism for hydrated proton mobility in liquid water.* The Journal of Chemical Physics, 2005. **122**: 014506.

23. Dellago, C. and G. Hummer, *Kinetics and Mechanism of Proton Transport across Membrane Nanopores.* Physical Review Letters, 2006. **97**: 245901.

24. Dellago, C., M. M. Naor, and G. Hummer, *Proton Transport through Water-Filled Carbon Nanotubes.* Physical Review Letters, 2003. **90**: 105902.

25. Pomes, R. and B. Roux, *Theoretical Study of H+ Translocation along a Model Proton Wire.* Journal of Physical Chemistry, 1996. **100**: p. 2519-2527.

26. Pomes, R. and B. Roux, *Free Energy Profiles for H+ Conduction along Hydrogen-Bonded Chains of Water Molecules.* Biophysical Journal, 1998. **75**: p. 33-40.

27. Wu, Y. and G.A. Voth, *A Computer Simulation Study of the Hydrated Proton in a Synthetic Proton Channel.* Biophysical Journal, 2003. **85**: p. 864-875.

28. VandeVondele, J., et al., *The influence of temperature and density functional models in ab initio molecular dynamics simulation of liquid water.* The Journal of Chemical Physics, 2005. **122**(1): 014515.

29. Lee, H.-S. and M.E. Tuckerman, *Structure of liquid water at ambient temperature from ab initio molecular dynamics performed in the complete basis set limit.* The Journal of Chemical Physics, 2006. **125**(15): 154507.

30. Hura, G., et al., *A high-quality x-ray scattering experiment on liquid water at ambient conditions.* The Journal of Chemical Physics, 2000. **113**(20): p. 9140-9148.

31. Svozil, D. and P. Jungwirth, *Cluster Model for the Ionic Product of Water: Accuracy and Limitations of Common Density Functional Methods.* The Journal of Physical Chemistry A, 2006. **110**(29): p. 9194-9199.

Chapter 4:

The Poly-glycine-serine and Poly-glycine-aspartic Models

The addition of polar serine and aspartic side-chain residues to the poly-glycine scaffold and their effect on water behavior and the proton transport process

4.1. Introduction

The work in this Chapter moves on from the baseline poly-GLY model introduced in Chapter 3. Two further models are presented, both based on a modified poly-GLY scaffold. In the first, two oxidisable serine residues (SER, side chain R = -CH_2OH), have been added to the poly-glycine scaffold in order to create a model which can interact directly with the water molecules in the central cavity. In the second model, three aspartic residues (ASP, side chain R = -COOH) were added to the poly-glycine helix. Into the central cavities ten and six water molecules were added to the two models, respectively, along with an excess proton. Both models therefore have a permanent water wire. As with the poly-glycine model presented in the previous chapter, there is no chemical driver for the PT process and thus there is no bias on the direction of movement of the proton.

The key questions to address in this Chapter are:

- Is the mechanism for PT in these channels consistent with that observed for the poly-GLY model?
- Do the serine and/or aspartic residues participate directly in proton transport?
- How do the water molecules bind to the polar serine and aspartic residues in comparison with the non-interacting glycine H-side chains? Does their presence 'fix' the positions of the water molecules in the channel?

4.2. Simulation Methods

Models: A polypeptide chain of glycine was twisted into a right handed alpha helix (with standard Phi and Psi angles of -57.0° and -47.0°, respectively) using the ARGUSLAB [1] polypeptide builder to create the basic scaffold. To construct the poly-GLY-SER model, two serine (SER) side chain R-groups (-CH_2OH) were then incorporated into the helical

backbone using the Materials Studio suite of software. The poly-GLY-ASP model was created by adding three equally spaced aspartic (ASP) residues along the poly-glycine helix backbone. Periodic boundary conditions (PBCs) were then imposed by placing

one full repeat of the helix (*i.e.* four helical turns, comprising fifteen amino acid residues) at the corner of an orthorhombic cell to lie along cell vector *b*, corresponding to 26.36 Å (see Figure 4.1). Thus cell vector *a* (= *c*, initially set to 8.0 Å and 10.5 Å for the poly-GLY-SER and -ASP respectively) defines the diameter of the inter-helical pore, into which the water environment is placed [see Figure 4.1(a) and (b)]. In this way the one helix-model is replicated to four, and in turn this packing arrangement is repeated to infinity along the *a,b* and *c* vectors. The central cavities were then 'filled' with water molecules, with ten molecules chosen for poly-GLY-SER and six for poly-GLY-ASP. These numbers were chosen so that a continuous interacting water/amino acid chain was created that ran the length of the *b*-axes for the two unit cells.

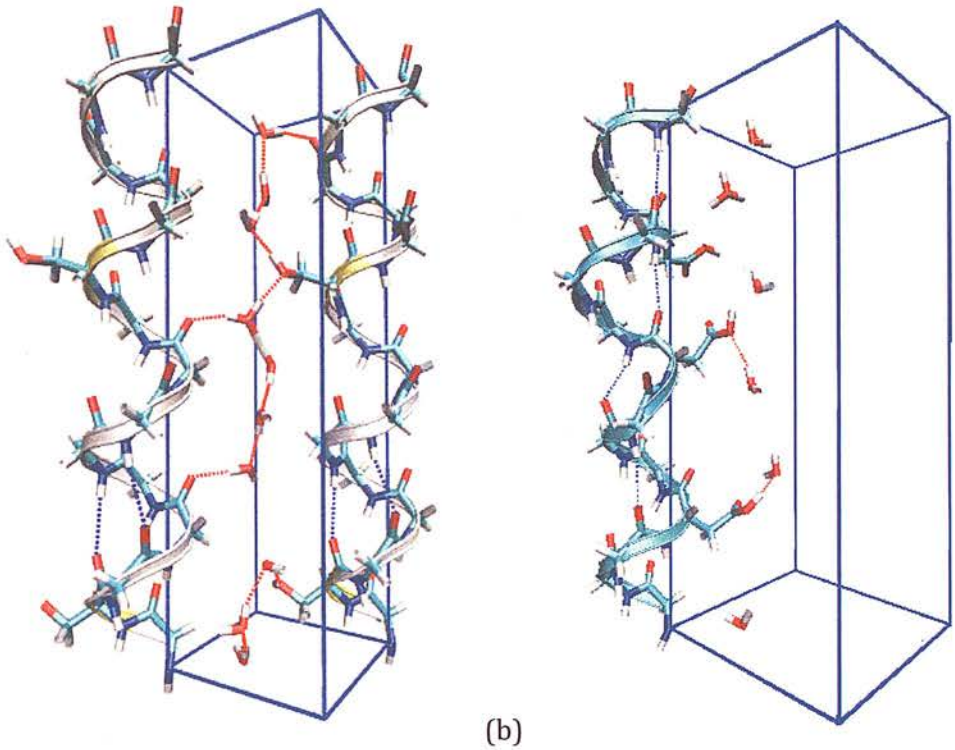
Simulations: All simulations were run using the CP2K[2,3] molecular dynamics simulation package with the GGA Becke-Lee-Yang-Parr (BLYP)[4,5] functional, coupled to a dual localised (Gaussian) and plane-wave basis set description. The localised basis set was of double-zeta quality and optimised for use against the Goedecker-Teter-Hutter set of pseudopotentials[6] (incorporated to model the core electrons) coupled to the BLYP functional. A series of single-point energy calculations determined the optimum energy cut-off (300Ry), and the subsequent geometry optimisation was performed in two steps using the Broyden-Fletcher-Goldfarb-Shanno (BFGS)[7-10] method. In the first instance atom positions only were allowed to relax; this was then followed by a series of single-point energy calculations to obtain the optimised cells vectors (as shown in Table 4.1)

Table 4.1. Optimised cell vectors for the poly-GLY-SER and poly-GLY-ASP models

Vector	Poly-GLY-SER, Å	Poly-GLY-ASP, Å
a,c	7.95	10.5
b	26.36	26.26

Dispersion, or van der Waals, interactions, which are inherently missing from DFT calculations were accounted for by the addition of a pair potential as suggested by Grimme.[11] Positional restraints were placed on every backbone nitrogen atom (fifteen in total) in the poly-GLY-SER model, but only every other nitrogen for the poly-GLY-ASP model. NVT MD simulations (maintained at 365K by a chain of Nose-Hoover

thermostats) were run using the same basis set as described above for *ca.* 30 ps, advancing in time increment steps of 0.55 fs. Visualisation of the models and the MD trajectories was performed using the VMD package.[12]



(a) (b)

Figure 4.1 (a) Poly-glycine-serine simulation cell and (b) Poly-glycine-aspartic simulation cell. The second helix has been included in the poly-GLY-SER cell to show the extent of the hydrogen bonding with a second helix in the cell. The extra width of the poly-GLY-ASP simulation cell means that this level of hydrogen bonding is not possible. Note the numbering of water molecules in the text is defined with respect to the bottom of both cells.

4.3. Results and Discussion

4.3.1. Structural Stability of the model

In this study both models used are based on a single alpha helix and water chain under full PBCs with added dispersion corrections and atomic positional

restraints placed either on every nitrogen atom of the backbone for poly-GLY-SER or every other atom for poly-GLY-ASP. The rest of the helix has completely free movement and thus the stability of the model must be checked. As with the polyglycine model presented in Chapter 3, the integrity of the PBCs was first investigated by verifying that the time-averaged C-N bond (which forms the continuous helix by bridging the cell boundary) is consistent with the time-averaged C-N bonds present in the helix backbone. A comparison of the boundary value with three other C-N bonds found they were within two standard deviations of an integral helix bond and thus confirmed that the boundary conditions remain intact and both channels can be modelled as continuous.

Figure 4.2 shows the root mean square deviation (rmsd) of the helix backbone over the course of the 30 ps NVT run relative to the initial (t = 0, *i.e.* optimised) structure. There is little fluctuation in atom position in the poly-GLY-SER model [Figure 4.2(a)], with an average displacement of *ca.* 0.17(2) Å and a maximum of 0.24 Å. The use of positional restraints on the nitrogen atoms of the helix has reduced the extent of helix movement compared to the unconstrained poly-glycine helix of the previous chapter and that of similar model systems.[15] Most helix movement occurs in the first picosecond of the simulation, after which the plot levels off, indicating that the helix does not breakdown during the course of the simulation.

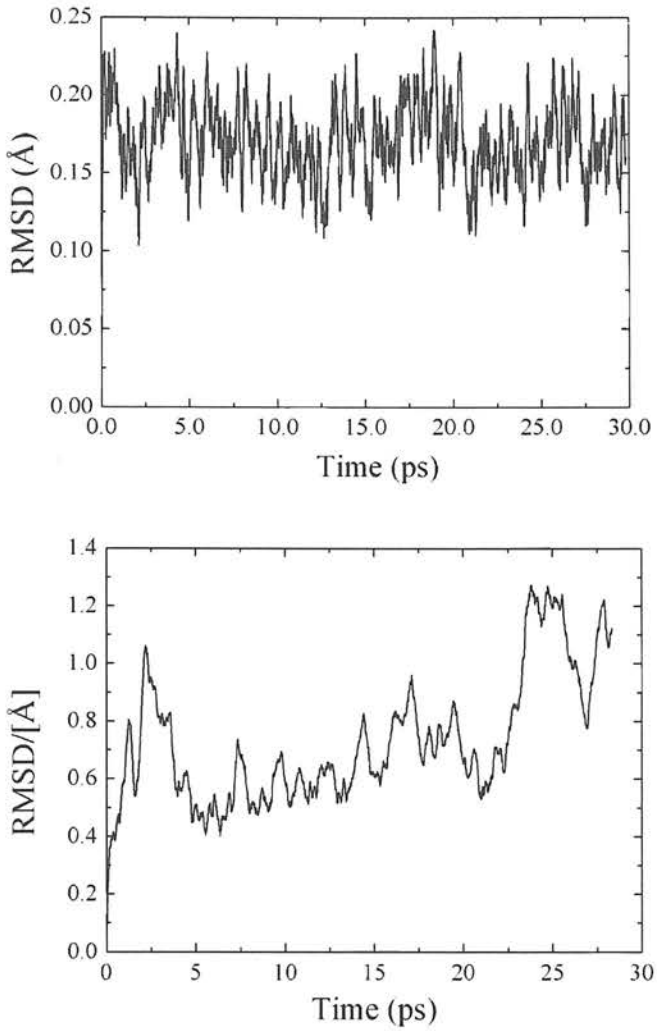
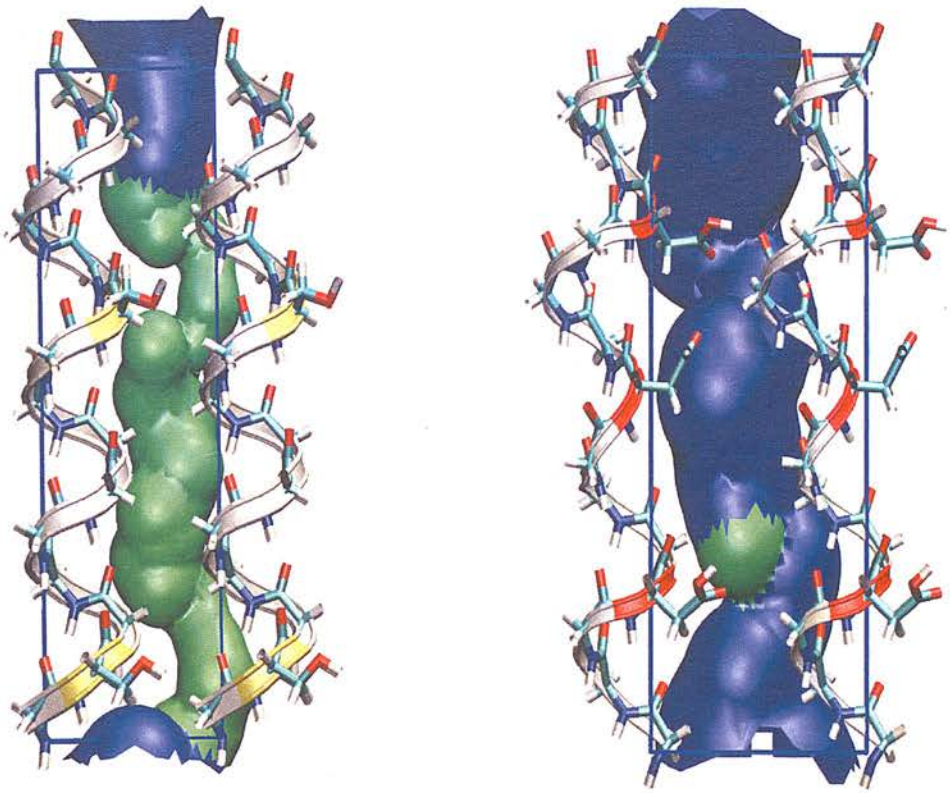


Figure 4.2 Relative Mean Square Deviation of the helix backbone of (a) poly-GLY-SER and (b) poly-GLY-ASP

Figure 4.2 (b) indicates increased fluctuation of atom position in the poly-GLY-ASP model in comparison with the poly-SER model, with an average displacement of *ca.* 0.718(2) (maximum 1.27 Å). The use of positional restraints just on every other nitrogen atom of the helix, coupled with more movement of the larger aspartic residues has allowed for increased helix backbone movement in relation to the poly-SER model discussed above. There is considerable movement in the first few picoseconds of the simulation, this then levels off for approximately 20 ps and then jumps again. Scrutiny of the helix structure at this point does indicate that the structure of the helix may not be as stable as in prior models, but RMSD remains at comparative levels for the duration of the simulation.

The channel radii as a function of the *b*-axis was further investigated using the HOLE program, [13] which maps the internal cavity of the channel based on the maximum radii of the hard spheres. Figure 4.3 shows the resulting representations, where the green regions denote a radius in the region 1.5-2.3 Å, and blue a wider range of $r > 2.3$ Å. This is well in excess of the minimum radius required for water intrusion (1.15 Å), and therefore indicates that all parts of the channel are accessible to water molecules.



(a) (b)

Figure 4.3 HOLE plot of the simulation cell detailing the internal cavity of the (a) poly-GLY-SER and (b) poly-GLY-ASP channels

4.3.2. The effect of serine residues on water-wire structure and dynamics

A time-averaged plot obtained for the poly-GLY-SER model over the 30 ps NVT trajectory is given in Figure 4.4 and shows that the excess proton is delocalised within a water-wire consisting of a maximum of five water molecules. The positioning of the

water molecules in the time averaged plot, in fact, shows little variation from what was obtained in the geometry optimisation process. Or in other words there is very little water movement in this model. The five water molecules that bear the excess proton have become trapped between the two -OH groups and are therefore separated from the rest of the water molecules in the channel. This naturally limits the distance that the proton can travel to around 8 Å. At no point was proton transfer observed to occur through the SER R-groups. These residues have therefore created a break in the chain of water molecules which the excess proton cannot bridge.

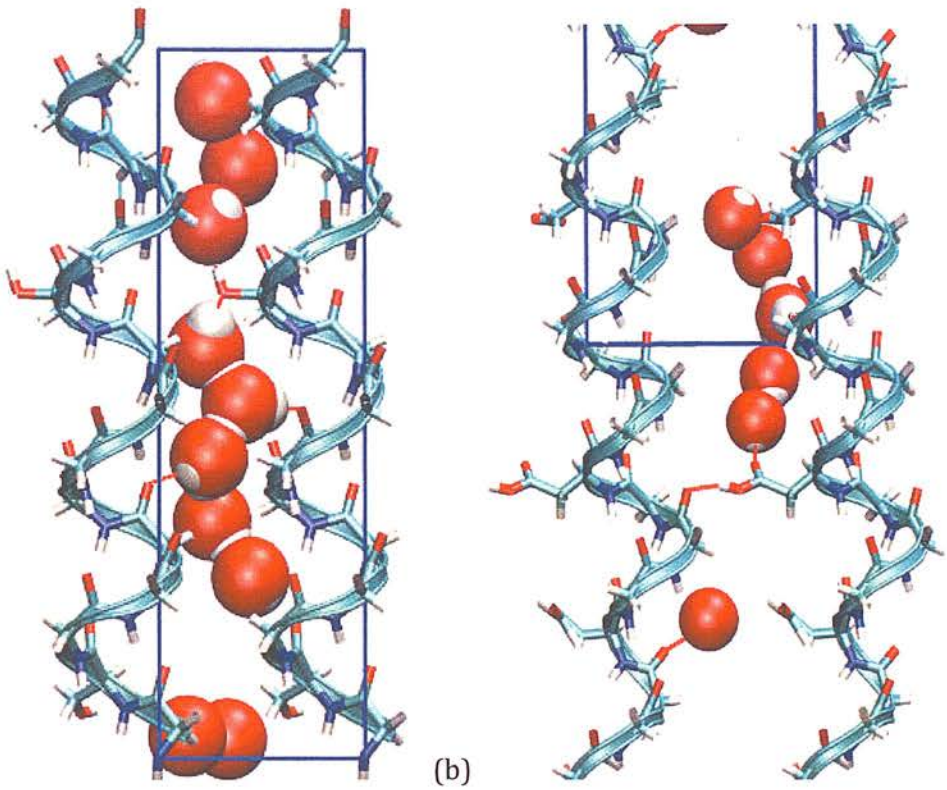


Figure 4.4 Time-averaged structures of the (a) ten water molecule poly-GLY-SER model and (b) the six water molecule poly-GLY-ASP model.

The average atomic distributions of the ten water molecules, two SER and three helix carbonyl oxygen-atoms have been used to further investigate the effect of adding serine residues on the behaviour of the water molecules along the channel *b*-axis. Figure 4.5 shows the peak distributions for the water molecules, along with the SER oxygen atoms

and those carbonyl oxygen atoms that are most strongly interacting with the central water wire.

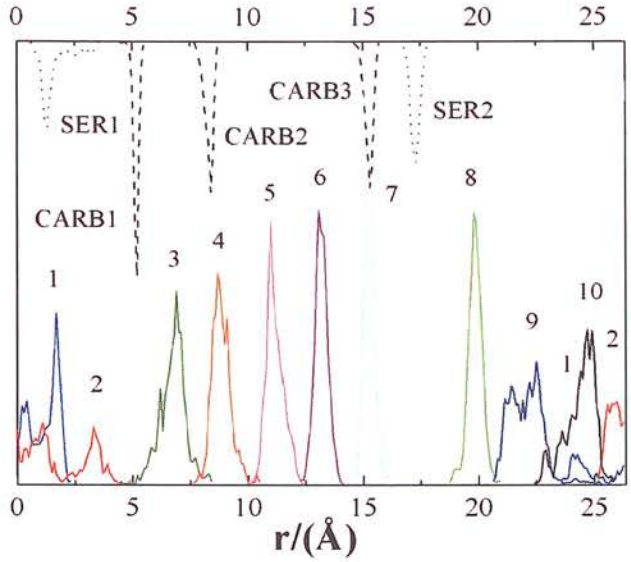


Figure 4.5 Average atomic distributions of water oxygen (bottom axis, coloured lines) and helix serine O- and carbonyl O-atoms (top axis, black dotted and dashed lines respectively) modelled along the channel axis (see Figure 4.1 for numbering scheme).

The plot clearly shows two regions, with those water molecules which participate in PT easily distinguished from those which do not. Narrow peak distributions, as evidenced for $H_2O(3\text{-}7)$ indicate strong hydrogen bonding interactions and the presence of a coordinated water-wire. The excess proton travels over four of these five water molecules over a distance of approximately 8.0 Å, a quarter of the channel length, without any significant diffusion of the water molecules themselves. $H_2O(8)$ remains tightly coordinated to SER2 throughout the simulation and as such also has a narrow peak distribution, whereas the wider peaks for molecules 1, 2, 9 and 10 indicate an increased mobility as compared to those molecules within the water wire.

4.3.3. The effect of aspartic residues on water-wire structure and dynamics

A time-averaged plot of poly-GLY-ASP is given in Figure 4.4(a) and shows that five of the six water molecules become 'trapped' between ASP3 and ASP1, whereas the sixth molecule remains co-ordinated to ASP2. The excess proton is delocalised within a wire constructed from these five water molecules, the two ASP R-groups and one carbonyl oxygen. Note the wire straddles the PBC boundary, and so the unit cell has been redrawn in Figure 4.6(b) to ease visualisation. The excess proton does not travel far from its starting position on $H_2O(5)$. PT is observed from $H_2O(5)$ to $H_2O(6)$ with the proton coming to rest on $H_2O(1)$. The positioning of the excess proton on $H_2O(1)$ leads to the formation of an Eigen-type complex. This complex traditionally consists of an oxonium ion plus three additional water molecules to form the $[H_9O_4]^+$ motif. In this case, however, one of the water molecules has been replaced by a carbonyl oxygen. In fact, during the simulation, the proton spends approximately 48% of its time directly co-ordinated to this carbonyl oxygen forming a structure more in line with a Zundel or $[H_5O_2]^+$ where the proton is shared equally between two oxygens.

The presence of the Eigen-type cation has, in previous work, been linked to the formation of a 'proton-trap' which can limit PT [14]; the proton can now travel along either of the x- or y-axes, whereas in a narrow channel it only has the possibility of the y-axis.

The average atomic distributions of the six water molecules and three ASP residues have been used to see the effect of adding ASP residues on the behaviour of the water molecules along the channel b-axis. Figure 4.6 shows the peak distributions for the water molecules, along with the aspartic oxygen atoms. The five water molecules [$H_2O(5,6$ and $1\text{-}3)$] that form a sustained complex do have some movement, as shown below, but the majority of this movement occurs at the start of the MD trajectory. $H_2O(5)$ in particular is held in place by ASP3 for the majority of the simulation, as shown by the relatively narrow distribution.

$H_2O(4)$ remains co-ordinated to a carbonyl backbone oxygen and also ASP2 for the duration of the simulation and as such is excluded from the water-wire.

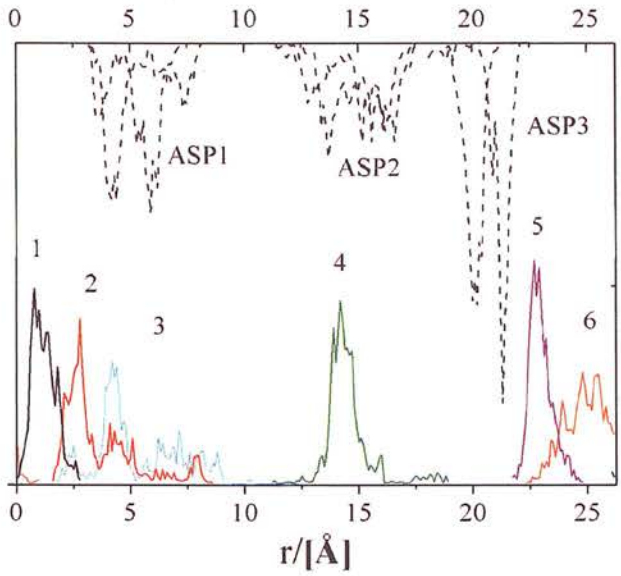


Figure 4.6 Average atomic distributions of water oxygen (bottom axis, coloured lines) and helix serine O- and carbonyl O-atoms (top axis, black dotted and dashed lines respectively) modelled along the channel axis.

4.3.4. Extent of delocalisation of the excess proton

The average lifetimes of cationic states created during the MD trajectories, and the instances that those geometric criteria for those states were fulfilled were calculated in the same way as for the polyglycine model (see Section 3.4.4). The results, along with the lifetimes, are given in Table 4.2, alongside the results obtained for the polyglycine model for direct comparison. A wide range of cationic state lifetimes are observed, from tens of femtoseconds, to several picoseconds. There is a rough correlation between the lifetimes and the prevalences for each state, that is in general the more times the state is formed, the longer is its lifetime. In common with the poly-glycine model, at no point was an isolated Zundel (n = 2) cation observed, however the n > 3 states may contain the Zundel structure solvated by n – 2 water molecules. All results are consistent with the time-averaged plots in Figures 4.4 and 4.6 and the average water distributions in Figure 4.5 and 4.7.

Table 4.2 Prevalences (expressed as a % of the total simulation time) and Lifetimes (in fs) of the cationic states $[H_{2n+1}O_n]^+$ observed during the simulations.

	Poly-GLY-SER		**Poly-GLY**		**Poly-GLY-ASP**	
n	*Prevalence (%)*	*Lifetime (fs)*	*Prevalence (%)*	*Lifetime (fs)*	*Prevalence (%)*	*Lifetime (fs)*
2	0.1	27	0	0	30.6	4221
3	65.8	3205	0.5	9	9.9	377
4	7.7	1224	43.7	2569	5.8	204
5	26.2	2149	4.7	53	10.4	224
6	0	0	2.2	22	43.3	6782
7	0	0	12.3	942	0	0
8	0	0	18.6	210	0	0
9	0	0	18	517	0	0
10	0	0	0.1	10	0	0

4.3.4.1. The Poly-glycine-serine model

The ten water-molecule poly-GLY-SER model supports the formation of water-wires from n = 2-5 with no possibility of wires of longer length. This is due to the influence of the SER groups, which isolate the excess proton and five water molecules from the rest of the channel. This is the first major difference observed for this model compared to the baseline poly-glycine model, where wires of up to n = 9 are observed throughout the course of the simulation. This model system shows a preference for the n = 3 state, (*i.e.* $[H_7O_3]^+$), with a prevalence of 65.8 % and a lifetime of 3.2 ps. Wires of length n = 4 and n = 5 are also formed with significant frequency, whereas a wire containing only two water molecules is rarely observed and has a lifetime of only 27 fs, which is on a similar timescale to an O-H bond stretching vibration.

As observed in the time-averaged plot, the water molecules form consistent hydrogen bonds with both helix backbone carbonyl oxygen atoms CARB1, CARB2 and CARB3, and also with the oxygen atom of SER2. The above method was therefore

extended to investigate the extent of delocalisation over these functional groups. The criteria for successful hydrogen-bonding was the same as used for the water molecules alone. Again, a hydrogen bond was defined by considering interactions between oxygen atoms that fell within a cut-off distance of $R_{OO} \leq 3.5$ Å, along with an H-O...O angle constraint of 30°. The results from this second cationic lifetime study are presented in Table 4.3. The inclusion of these four extra groups allows for delocalisation over a maximum of nine oxygen atoms, and a clear preference is exhibited for a 'wire' which now includes five oxygen atoms. On closer inspection, however, this five-membered wire has its main contribution from a complex of three water molecules coordinated to SER1 and CARB3. This is consistent with the results presented in Table 4.2 which exclude any specific helical interactions.

Table 4.3 Prevalences (expressed as a % of the total simulation time) and lifetimes (in fs) of the cationic states in Table 4.2 for the poly-GLY-SER model (where n = number of oxygen atoms) extended to include helical interactions.

	Poly-GLY-SER		**Poly-GLY-ASP**	
n	*Prevalence (%)*	*Lifetime (fs)*	*Prevalence (%)*	*Lifetime (fs)*
2	0.02	44	7	55
3	0.1	20	22	340
4	4.7	68	6	188
5	61.2	2468	28	3196
6	0.4	6.3	2	72
7	8.4	735	8	272
8	13.3	602	16	2602
9	11.7	716	22	2965

4.3.4.2. The Poly-glycine-aspartic model

Table 4.2 shows the lifetimes and prevalence of cationic states formed in the poly-GLY-ASP model when all helical interactions bar that of one carbonyl oxygen have been excluded. The carbonyl has been included in this case as a sustained hydrogen bond is formed between it and the water holding the excess proton which lasts for the

duration of the simulation (in fact the excess proton is directly co-ordinated to this carbonyl oxygen for 48% of the simulation).

If only these interactions are considered a complex containing six oxygen atoms is most prevalent during the simulation, which agrees to the time-average structure in Figure 4.6. The second most prevalent complex is that of a Zundel which is normally classed as two water molecules with a central excess proton forming the $[H_5O_2]^+$ motif which has been discussed at length in the introductory chapter. In this case, however the complex consists of a glycine carbonyl oxygen (CARB1) and a water molecule with excess proton. In fact, the excess proton spends almost half the simulation (48%) resting on the CARB1 with a hydrogen bond to $H_20(1)$ (as evidenced in Figure 4.8 below).

When the analysis is extended to include all helical interactions a clear preference for the formation of longer sustained eight- and nine-membered water-wires is exhibited. These are formed by the addition of two aspartic oxygens (ASP3 and ASP1) and a further helix carbonyl oxygen, which allows for increased delocalisation of the excess proton as shown in Figure 4.8 below. Note: Those carbonyl oxygens which are not directly hydrogen-bonded to the water-wire have not been included in the analysis. The two aspartic acid head groups form a barrier to PT by co-ordinating the water molecules to the area of the channel between them, meaning the excess proton cannot travel beyond these groups. They themselves do not participate directly in PT during the simulation.

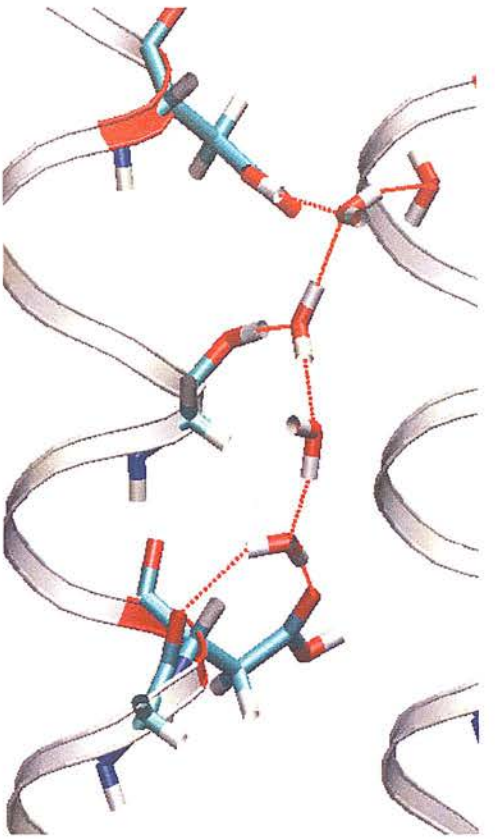


Figure 4.7 The nine-oxygen complex which is most prevalent in the poly-GLY-ASP model system. Note: The carbonyl oxygen atoms not directly H-bonded have not been included in the numbering of the complex.

4.3.5. The mechanism for proton transport

The proton travels spontaneously over four water molecules in the poly-GLY-SER channel, covering a distance of approximately 8 Å. The mechanism observed is, as expected, consistent with that obtained in the poly-GLY model presented in the previous Chapter. For completeness, the time-dependent RDFs centred on the oxygen which will accept the proton (O_{next}) and its other nearest neighbour ($O_{nearest}$) averaged over the whole simulation are also presented here for the poly-GLY-SER model (see Figure 4.8).

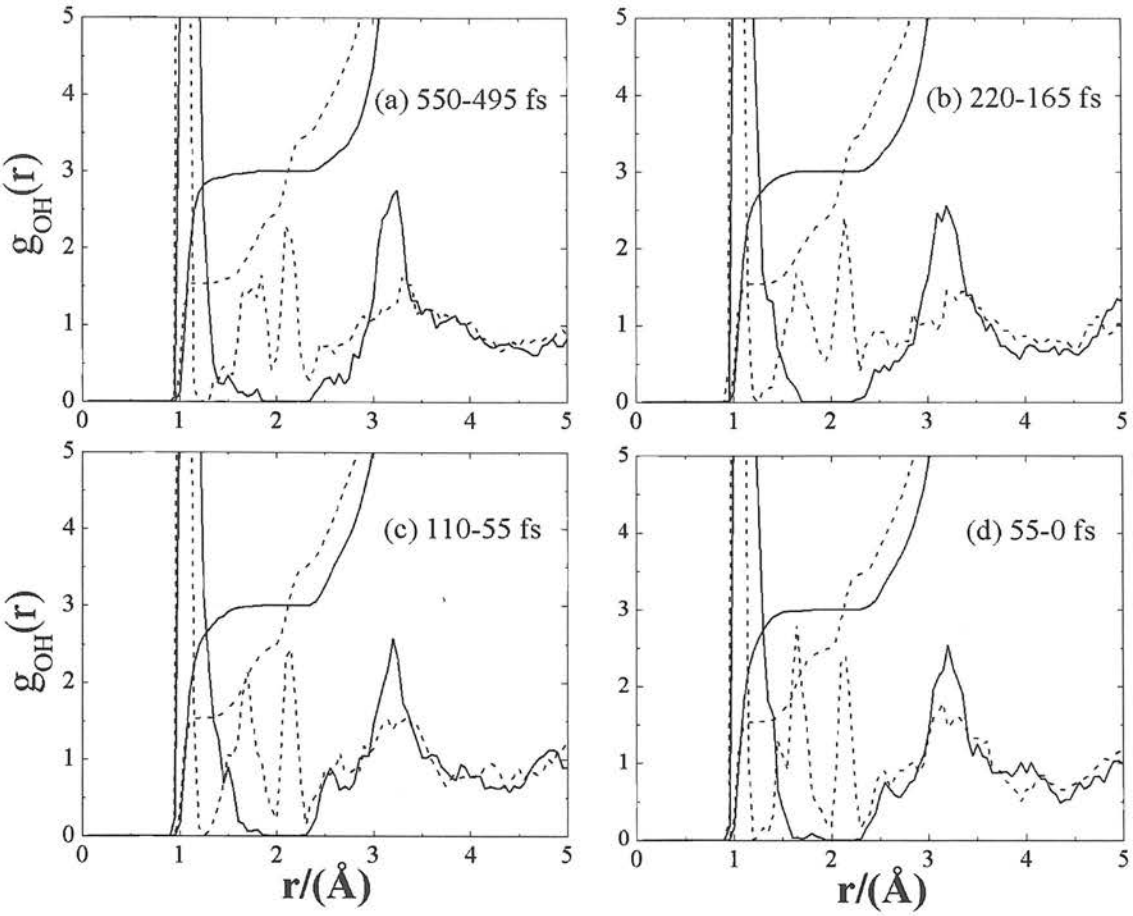


Figure 4.8 Time dependent RDFs and integrated coordination number plots of O_{next} (solid black line) and $O_{nearesr}$ (dashed black line) averaged over 55 fs intervals before [(a) 550-495 fs] and approaching a PT event [(b) 220-165 fs, (c) 110-55 fs and (d) 55-0 fs].

Starting with the $O_{nearest}$ g(r) plot (dashed black line, Figure 4.8), four peaks are observed, namely,

i. the dominant peak at 1.1 Å, which is assigned to the hydrogen atoms directly bonded to the oxygen atoms to form a water molecule or a serine O-H group,
ii. a broad peak centred at *ca.* 1.45 Å which can be attributed to an O...H hydrogen bond interaction between a SER oxygen and a backbone hydrogen atom,
iii. a sharp peak centred at 1.7 Å which corresponds to O...H interactions between water molecules,

iv. and finally a very broad peak at around 3.3 Å which is an O...H distance over a wider coordination sphere.

This plot remains essentially constant throughout the four intervals, with only small deviations in position. Peaks one, three and four are consistent with the findings for poly-glycine and bulk water, and the addition of the second peak is easily attributed to the serine residue. This indicates, that as in the poly-GLY model, once more the $H_2O_{nearest}$ molecule does not play a role in the PT mechanism.

The O_{next} plot (solid black line, Figure 4.8), does however show significant changes over the time intervals shown. In this case there are only three peaks. The excess proton is never successfully transferred to a serine residue, which therefore never takes on the role of O_{next}. As such there is no incidence of an O...H separation corresponding to a serine hydrogen bond interaction and therefore there are no additional peaks over that observed for the poly-glycine model. The dominant peak is again centred around 1.1 Å, the second peak now appears as a shoulder at *ca.* 1.5 Å and the third peak is now sharper and at a separation of 3.1 Å . Over the course of the PT process the second peak shoulder shifts to the left (from 1.5 Å to 1.35 Å) indicating the formation of the short, strong hydrogen bond of the Zundel cation, $[H_2O_5^+]$, an important precursor to successful PT. These findings are again in agreement with previous findings for both bulk water and the poly-glycine channel.

In Chapter 3 two important differences between PT in bulk water and in a channel environment were observed. First, in a channel environment all successful PT steps are preceded by a large number of unsuccessful or 'rattling' PT events, where the next proton transfer step sees the proton simply reverse its direction. This phenomenon does not occur in simulations of the excess proton in bulk water. [15] Figure 4.9 shows the identity of the oxonium ion (O*) *versus* time for the whole of the trajectory. Again this plot is in agreement with the previous findings for the poly-glycine channel: the rattling events are evident from the large number of incidences where the O* identity repeatedly flips with a neighbour. Again the direction of PT is limited in the channel environment as compared with bulk water and the positions of the water molecules are constrained by the presence of the helix and helical side chain groups.

The second difference observed in these simulations compared to those reported for bulk water is that there is no change in the coordination number of O_{next} over the course of the PT process, it remains constant at three. In bulk water O_{next} loses one hydrogen bond as a result of PT, with its coordination changing from four to three. In the channel environment there is a limit on the number of hydrogen bonds a water molecule can form. In both the poly-GLY and the poly-GLY-SER channel a maximum coordination of three is observed, *i.e.* two hydrogen atoms directly coordinated to the central oxygen atom along with a third hydrogen from an accepted hydrogen bond. This limit on the number of hydrogen bond interactions in the channel environment means that, unlike the bulk water mechanism, losing a hydrogen bond is now no longer a necessary precursor to successful PT.

Continuing the comparison with both the poly-GLY channel and bulk water, we now look at the behaviour of $O_{next}/O_{nearest}$ in the run up to a PT step. A 'special pair dance' has been observed in the bulk water system whereby the identity of O_{next} and $O_{nearest}$ fluctuates as they vie for position to receive the excess proton. In the poly-GLY channel environment it was not possible to definitively comment as to whether we see the same phenomenon due to the increased incidence of rattling events. These rattling events are again present in the poly-GLY-SER channel [see Figure 4.9(a) for O* identity and (b) for O_{next} identity] as discussed above. Figures 4.9 (c) and (d) focus on a 0.5 ps window of the simulation. There are some short periods where the identity of the O* does not change, and at these points there is some indication of O_{next} and $O_{nearest}$ identity flipping [see Figure 4.7(d)]. There are not enough incidences of stationary O* to state with certainty whether or not the same 'special pair dance' is an important part of the PT process in a channel environment, but the limited opportunity for the 'dance' to occur and the large number of PT steps observed would suggest not.

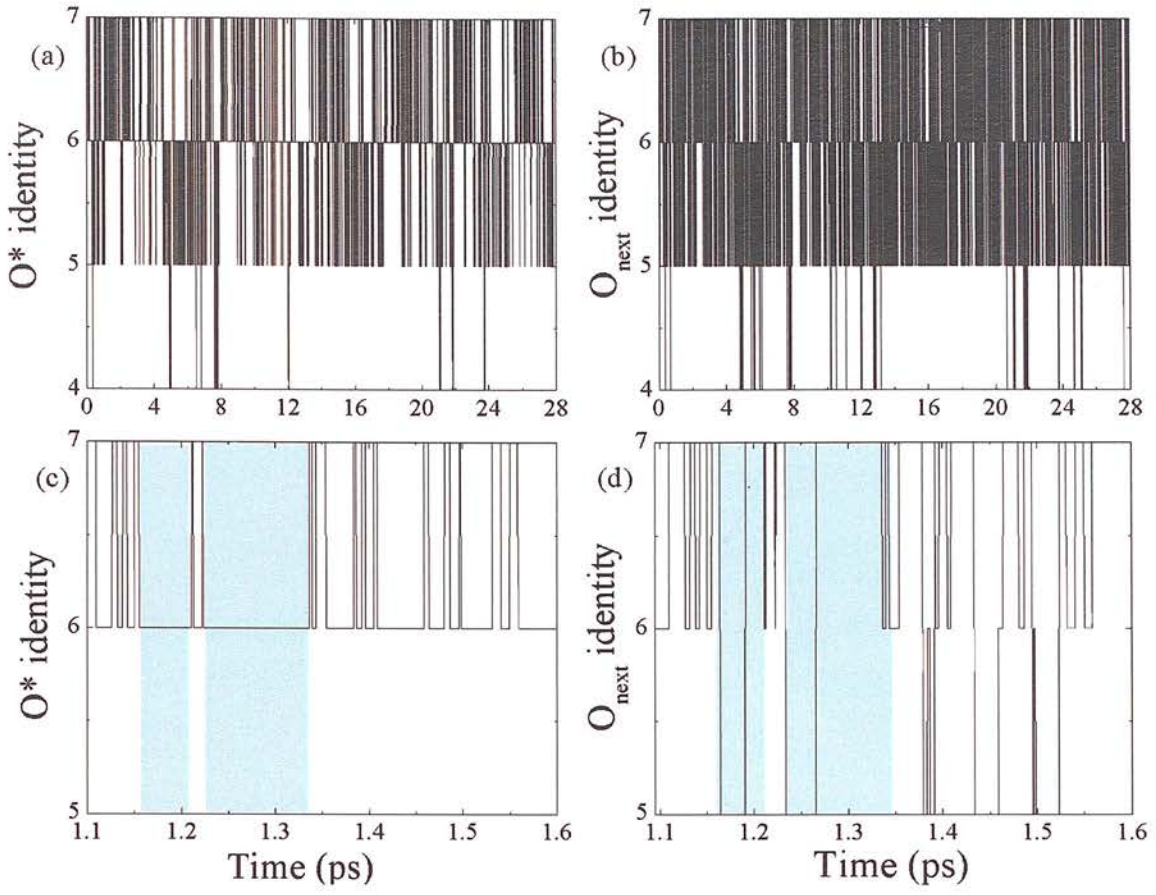


Figure 4.9 (a) Oxonium identity, (b) O_{next} identity for the entire MD simulation, (c) Oxonium identity and (d) O_{next} identity for a 0.5 ps window obtained in the poly-GLY-SER channel simulation.

It was not possible to perform time-dependent RDF analysis on the poly-GLY-ASP model as there was not a sufficient number of 'real' PT steps (i.e. those steps where the proton does not simply reverse its direction in the subsequent transfer step)

A plot of the identity of the O* and O_{next} for poly-GLY-ASP has been included below and may shed some light as to the occurrence of the 'special-pair dance'. The addition of the aspartic residues and subsequent widening of the channel allows for the formation of an Eigen-type cation (discussed previously) whereby the excess proton forms three hydrogen bonds; two with neighbouring water molecules and a third with a helix carbonyl group. The formation of this relatively stable Eigen-like complex means

that there are fewer PT steps and more prolonged periods resting on a single oxygen atom. This behaviour is in line with previous work where the Eigen structure was seen as a 'proton-trap' in a channel environment. [14] In addition, this is perhaps similar the environment immediately surrounding an excess proton in bulk water where the 'special-pair dance' has been observed.

Figure 4.10 (a) and (b) show the path of the excess proton over the entire MD simulation. The proton spends the majority of its time on the carbonyl oxygen (number one in the figure below) and then travels only over $H_2O(1)$ and $H_2O(2)$ (numbered two and three) for the remainder of the time. If we focus in one the behaviour of the system when the proton remains stationary for extended periods of time [highlighted sections in Figure 4.10(c) and (d)] flipping of the identity of the O_{next} can be clearly seen. This reinforces the reasoning that the narrow channel and lack of Eigen complex allows for both increased rattling and no special pair dance for the poly-GLY and poly-GLY-SER models previously discussed.

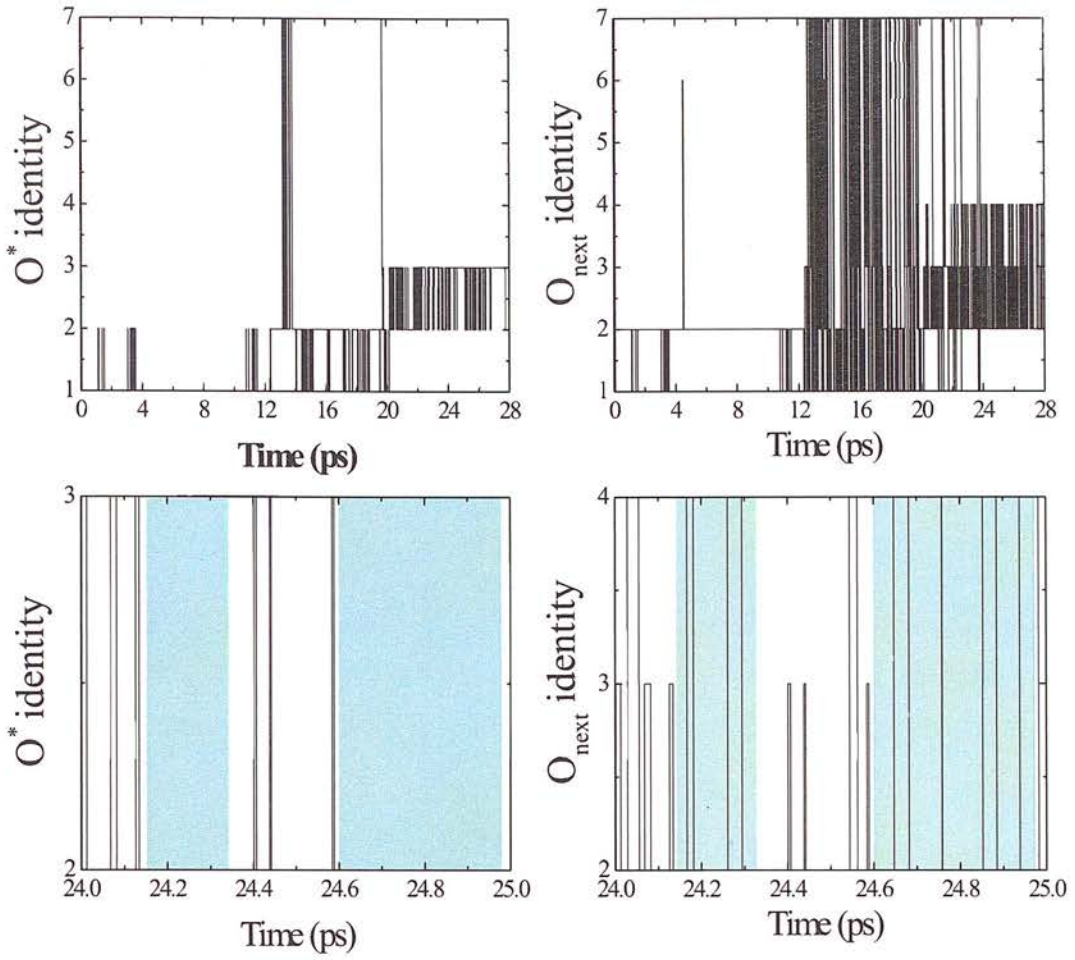


Figure 4.10 (a) Oxonium identity, (b) O_{next} identity for the entire MD simulation, (c) Oxonium identity and (d) O_{next} identity for a 0.5 ps window obtained in the poly-GLY-ASP channel simulation.

4.4 Conclusions

Two further models have been constructed by the addition of polar serine and aspartic residues to the poly-GLY scaffold discussed in Chapter 3. The polar –OH and CH_3COOH side chains are capable of direct interaction with the central water-wire. Fast proton transport has been observed in both models; the proton travels over a distance of 8 Å over four distinct model molecules in the poly-GLY-SER model, but in the poly-GLY-ASP model the distance is greatly reduced due to the formation of an Eigen-type cation.

Both polar residues exert a noticeable effect of the structure and dynamics of the associated water-wires, although neither are seen to participate directly in proton transport. In both cases five molecules become 'trapped' between the polar residues. In the case of the poly-GLY-SER model the influence is so great that their position does not change significantly throughout the course of the simulation. Lasting hydrogen bonds are formed with both the serine –OH groups and also with backbone carbonyl oxygen atoms. The SER residues therefore seem to provide a limit on the distance the proton can travel in this model. In the poly-GLY-ASP model there is considerably more movement in the model system as a whole, but the five membered water-wire again remains constant throughout the course of the simulation. The aspartic acid residues again provide a barrier to PT.

Subsequently the length of the water-wire is limited. Wires of length n = 2-9 are formed in the poly-GLY channel whereas, in this case, the maximum possible is a five-membered wire, with n = 3 the preferred state. An extension to the method shows that this n = 3 cationic species is part of a complex with includes two backbone carbonyl atoms. The proton is delocalised over five oxygen atoms in total.

Similar behaviour is observed in the poly-GLY-ASP model whereby a wire of length n = 6 is favoured if helical interactions (other than a carbonyl group) are ignored. This represents the five water molecules which are held between the aspartic acid side chain residues and one helix carbonyl oxygen. If the analysis is extended to include helical interactions the most prevalent state becomes that of a complex consisting of nine oxygen atoms (with the addition of two aspartic residue oxygens and a further helix carbonyl oxygen). The excess proton in this case spends approximately half the simulation (*ca.* 48%) co-ordinated to the helix carbonyl.

The mechanism for proton transport has been investigated using the poly-GLY-SER model and is consistent with that observed for the poly-GLY channel, the number of coordinated hydrogen atoms around an oxygen atom remains constant at three and there are a large number of rattling events. The lack of opportunity for the 'special-pair dance' to occur due to the increased rattling events, coupled with a large number of proton transfer events would suggest that this phenomenon is not an integral part of

the channel transfer process. Again the O* to O_{next} bond length shortens prior to proton transport, indicating the formation of a Zundel cation.

There were too few 'real' PT steps for TD-RDF analysis to be performed on the poly-GLY-ASP model. The formation of the Eigen-type cation means that there is opportunity for the proton to remain on a single oxygen atom for a prolonged period of time. Tracking the identity of the O_{next} while the identity of O^* remains constant has allowed for observation of the 'special-pair dance' which is part of the mechanism in bulk water. This reinforces the conclusion from the previous chapter that the narrow channel and increased rattling do not allow this phenomenon to occur. When the channel widens and the Eigen-type complex can be formed there is the option for PT in more than one plane and as such the 'special-pair dance' can be observed.

4.5. Bibliography

1. Thompson, M.A., *ArgusLab 4.0.1*, Planaria Software LLC: Seattle, WA.

2. Accelrys Inc.., *Materials Studio*. 2002: San Diego.

3. The CP2K Developers' Group. 2008, http://cp2k.berlios.de.

4. Becke, A.D., *Density-functional exchange-energy approximation with correct asymptotic behavior.* Physical Review A, 1988. **38**(6): p. 3098-3100

5. Lee, C., W. Yang, and R.G. Parr, *Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density.* Physical Review B, 1988. **37**: p. 785-789

6. Goedecker, S., M. Teter, and J. Hutter, *Separable dual-space Gaussian pseudopotentials.* Physical Review B (Condensed Matter), 1996. **54**(3): p. 1703-1710.

7. Broyden, C.G., *The Convergence of a Class of Double-rank Minimization Algorithms.* Journal of the Institute for Mathematics and Applications, 1970. **6**: p. 222-231.

8. Fletcher, R., *A new approach to variable metric algorithms.* Computer Journal,, 1970. **13**: p. 317-322.

9. Goldfarb, D., *A family of variable metric methods derived by variational means.* Mathematics of Computation, 1970. **24**: p. 23-36.

10. Shanno, D., *Conditioning of quasi-Newton methods for function minimization.* Mathematics of Computation, 1970. **24**: p. 647-656.

11. Grimme, S., *Semiempirical GGA-type density functional constructed with a long-range dispersion correction.* Journal of Computational Chemistry, 2006. **27**(15): p. 1787-1799.

12. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics.* Journal of Molecular Graphics, 1996. **14**: p. 33-38.

13. Smart, O.S., et al., *HOLE: A program for the analysis of the pore dimensions of ion channel structural models.* Journal of Molecular Graphics, 1996. **14**: p. 354-360.

14. Brewer, M.L., U.W. Schmitt, and G.A. Voth, *The Formation and Dynamics of Proton Wires in Channel Environments.* Biophysical Journal, 2001. **80**: p. 1691-1702

15. Shepherd, L.M.S. and C.A. Morrison, *Simulating Proton Transport through a Simplified Model for Trans-Membrane Proteins.* Journal of Physical Chemistry B, 2010. **114**: p. 7047-7055

Chapter 5:

The Poly-glycine-aspartic Model

A study into the preferred pathway of the excess proton

5.1 Introduction

The work presented in this Chapter moves on from the mechanistic studies performed in Chapters 3 and 4, and instead looks at the preferred pathway of an excess proton, and its iteraction with aspartic side chain residues. Three oxidisable aspartic acid residues (ASP, side chain R = CH_2COOH), have been added to the poly-GLY scaffold in order to create a model that can interact directly with the water molecules in the central cavity. Four models will be investigated based on water-wires of length 5, 6 and 9 molecules, and a water chain of 6 molecules. A chemical driver for the proton transfer (PT) reaction was created by removing a proton from the side-chain of one of the ASP residues and placing it on a water molecule some distance away to create an oxonium, $[H_3O]^+$ ion. This chemical driver results in the proton moving quickly back to its original site. Reprotonation occurs within the first 2 ps of a *ca.* 25 ps MD trajectory. Due to the rapid nature of the PT reaction, and the fact that computational constraints only allow for a single trajectory to be recorded, it must be noted that only one possible pathway per model was explored; many others must also exist. That said, the models do allow some general principles regarding PT within a mixed amino acid alpha-helical pore to be obtained.

The key questions to address in this Chapter are:

- Do the ASP residues participate directly in proton transport?
- How do water molecules bind to ASP residues?
- What is the effect of introducing a chemical gradient on the rate of PT?
- How does changing the density of water molecules affect the rate and path of PT and also the structure/movement of water molecules?

5.2 Simulation Methods

Model: A polypeptide chain of glycine was twisted into a right handed alpha helix (with standard Phi and Psi angles of -57.0° and -47.0°, respectively) using the ARGUSLAB[1] polypeptide builder. Three equally spaced aspartic acid (ASP) side chain

R-groups (-CH_2COOH) were then incorporated into the helical backbone using the Materials Studio suite of software. [2] The simulation was run under full periodic boundary conditions (PBCs) by placing one full repeat of the helix (*i.e.* four helical turns, comprising fifteen amino acid residues) at the corner of an orthorhombic cell to lie along cell vector *b*, corresponding to 26.36 Å (see Figure 5.1). Thus cell vector *a* (= *c*, initially set to 11.5 Å) defines the diameter of the inter-helical pore, into which the water environment (comprising five, six and nine water molecules) is placed (see Figure 5.2). In this way the one helix-model is replicated to four, and in turn this packing arrangement is repeated to infinity along the *a,b* and *c* vectors

Simulations: All simulations were run using the CP2K molecular dynamics simulation package [3] with the GGA Becke-Lee-Yang-Parr (BLYP) [4, 5] functional, coupled to a dual localised (Gaussian) and plane-wave basis set description. The localised basis set was of double-zeta quality and optimised for use against the Goedecker-Teter-Hutter [6] set of pseudopotentials (incorporated to model the core electrons) coupled to the BLYP functional. A series of single-point energy calculations determined the optimum energy cut-off (300 Ry), and the subsequent geometry optimisation was performed in two steps using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method. [7-10] In the first instance atom positions only were allowed to relax; this was then followed by a series of single-point energy calculations to obtain the optimised cells vectors, during which the *b* vector shortened slightly (to 26.26 Å) while the *a* and *c* vectors were reduced to 10.5 Å. The water-wire model was then constructed by removing a hydrogen atom from the middle aspartic acid residue (labelled 'ASP2') and adding it to one of the water molecules of the channel to create an oxonium ion, $[H_3O]^+$. The five and nine water-wire models were then built from this baseline model, by removing or adding further water molecules as necessary. Dispersion, or van der Waals, interactions, which are inherently missing from DFT calculations were accounted for by the addition of a pair potential as suggested by Grimme. [11] Positional restraints were placed on every other backbone nitrogen atom (eight in total). NVT MD simulations (maintained at 365K by a chain of Nose-Hoover thermostats) were run using the same basis set as described above for *ca.* 25 ps, advancing in time increment steps of 0.55 fs. Visualisation of the models and the MD trajectories was performed using the VMD package. [12]

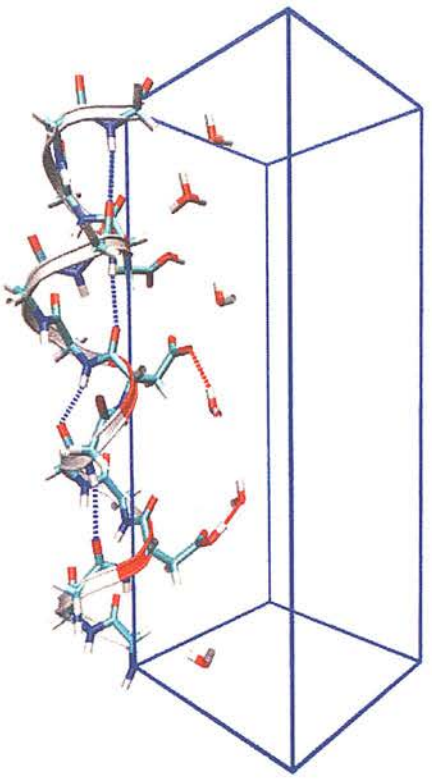
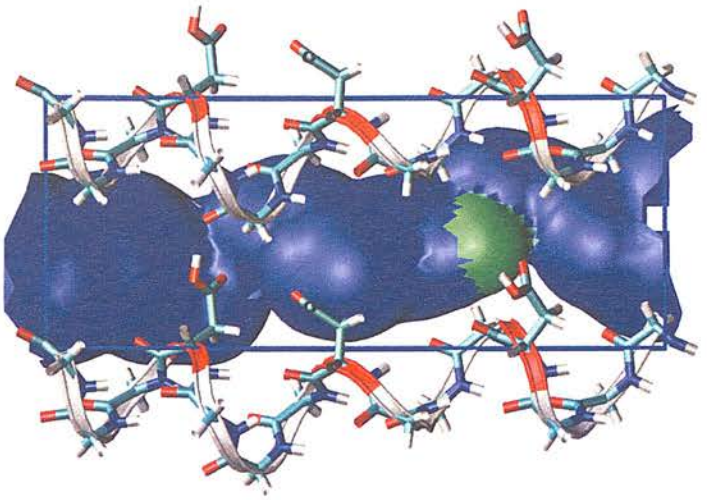
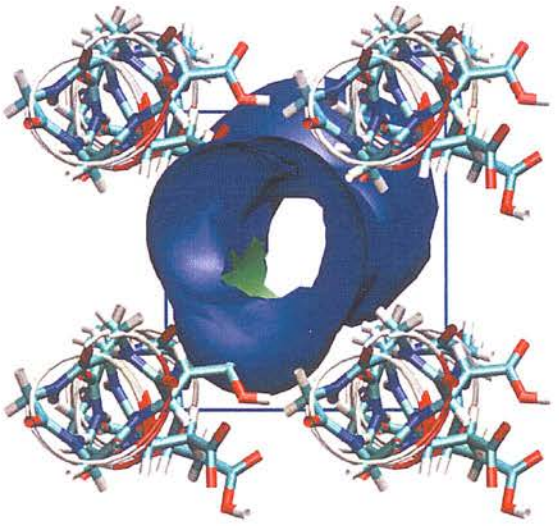


Figure 5.1 The six molecule water-wire simulation model with three ASP residues. Both water molecules and ASP residues are numbered sequentially from the bottom of the simulation cell.



(a)



(b)

Figure 5.2 HOLE representation of the internal cavity of the channel into which five, six or nine water molecules are placed.

5.3 Results and Discussion

5.3.1 Structural Stability of the model

In this study, the model used is a single alpha helix and water chain under full PBCs with added dispersion corrections and atomic positional restraints placed on every other nitrogen atom of the backbone. The rest of the helix has completely free movement and thus the stability of the model must be checked. The integrity of the PBCs was first investigated by verifying that the time-averaged C-N bond (which forms the continuous helix by bridging the cell boundary) is consistent with the time-averaged C-N bonds present in the helix backbone. A comparison of the boundary value with three other C-N bonds found they were identical to one standard deviation and confirms that the boundary conditions remain intact and the channel can be modelled as continuous.

Figure 5.3 shows the root mean square deviation (rmsd) of the helix backbone over the course of the 25 ps NVT run relative to the initial (t = 0, *i.e.* optimised) structure for all four of the model systems. The atom positions fluctuate noticeably, with average displacements of *ca.* 0.6(1) Å (and maximum values in the region of 0.8-1.0 Å). The use of positional restraints on the nitrogen atoms of the helix has reduced the extent of helix movement compared to the unconstrained poly-GLY helix of both our previous work and that of similar model systems. [13, 14] Most helix movement occurs in the first picosecond of the simulation, after which the plot levels off, indicating that the helix does not breakdown during the course of the simulation. The channel radius as a function of the *b*-axis was further investigated using the HOLE program [15], which maps the internal cavity of the channel based on the maximum radii of the hard spheres. Figure 5.2 shows the resulting representation, where the colour indicates the pore radius. Green denotes an intermediate range of 1.5 - 2.3 Å and blue a wider range of $r > 2.3$ Å. The available radius of the channel was observed to fluctuate around 2.5 Å (at the ASP residues) and 2.7 Å (at the GLY residues). This is well in excess of the minimum radius required for water intrusion (1.15 Å), and indicates that all parts of the channel are accessible to water molecules. Note the wider ends are

an artefact of the cavity being based on one simulation cell rather than the infinite channel created under PBCs.

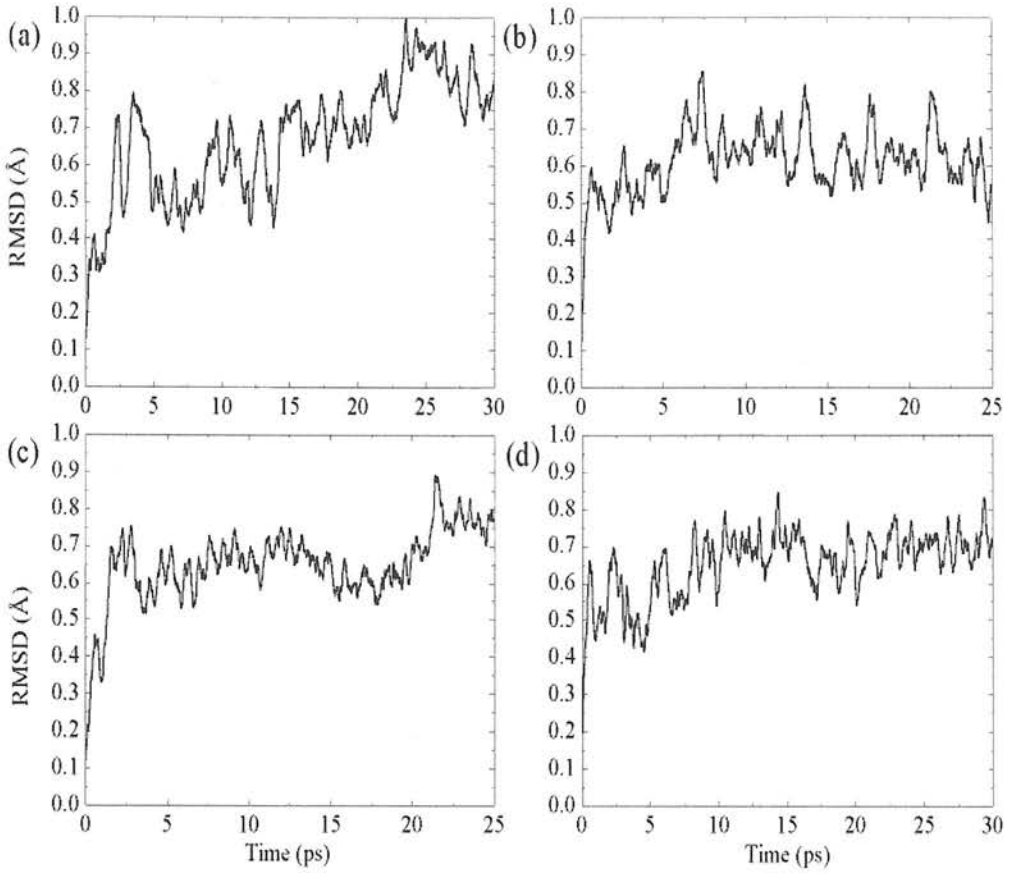


Figure 5.3 Relative mean square displacement (rmsd) of the helix backbone atoms for the (a) five water-wire model, (b) six water-wire model, (c) nine water-wire model and (d) six water-chain model.

5.3.2 Behaviour of the water wire vs. water chain

The effect of forming a water-wire by removing a proton from an ASP residue and placing it on a water molecule within the water-chain was investigated using oxygen-oxygen radial distribution functions (RDFs) or $g_{OO}(r)$ plots. These were

obtained for the six water molecule model where the proton has been moved (*i.e.* the water-wire model) and also for a similar model in which the proton was not moved (*i.e.* the water-chain model) These plots provide information not just on the nature of hydrogen bonding both within the water-wire and chain, but also between the water molecules and the alpha helix backbone, specifically, hydrogen- bonding to the ASP side chains and the helix backbone carbonyl groups.

Figure 5.4(a) shows the $g_{OO}(r)$ plots obtained for the two systems for the MD production run. The first (sharp) peak is identical for both models and occurs at a value of 2.35 Å. This can be assigned to the separation between the oxygen atoms of the ASP COOH group. The second (broad) peak of the distribution has a maximum at approximately 2.8 Å, with density observed from 2.55 Å, and is again consistent for both models. It can be assigned to a broad spectrum of O...O separations, including those between neighbouring water molecules and also between water molecules and helix carbonyl (CARB) and ASP oxygen atoms. This value of 2.8 Å is slightly longer than that recorded for the poly-GLY system (Figure 3.4) and can be explained in two ways. Firstly, there is a reduction in water molecules from ten to six, and secondly there is an increase in cell size. Together this results in an increased space for the water molecules, with the net effect being that the molecules display greater mobility than in the poly-GLY model. We note, however, that any effect of the excess proton is quickly lost in the data as the water-wire only exists for a very limited period of the simulation (*ca.* 8 % of the total simulation, representing just 2.2 ps).

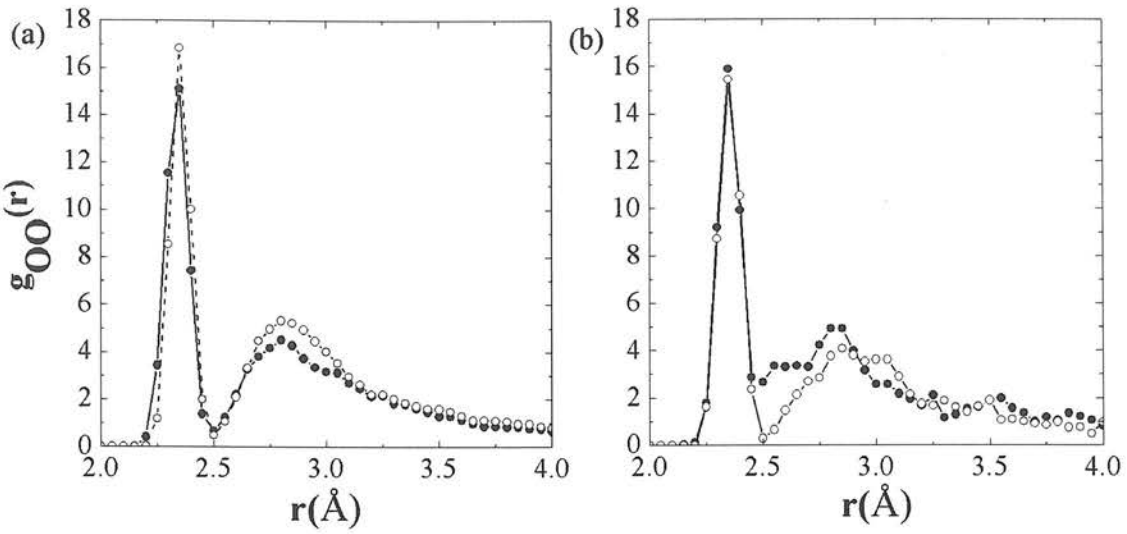


Figure 5.4 Radial distribution functions, $g_{OO}(r)$, obtained from (a) the six water-wire model (solid line) and the six water-chain model (dashed line) over the whole simulation (*ca* 55000 MD steps or 30 ps) and (b) similar plots taken over the first 4000 MD steps (2.2 ps).

In order to assess the effect of the excess proton on the dynamics of the water molecules, the above analysis must be repeated but using just the first 2.2 ps of the simulation. Figure 5.4 (b) is a $g_{OO}(r)$ plot of this time period. First it must be noted that, as expected, there is no change to the first peak. There is, however, some difference in the rest of the distribution. The water-chain model (*i.e.* neutral ASP residues) shows a much more cleanly defined first solvation shell structure, with the interactions between water molecules clearly starting at 2.5 Å; when the water-wire is present there is an increased density in the region 2.5-2.6 Å, suggesting that in the water-wire molecules can be positioned closer together. In previous work [14]these separations have been assigned to an Eigen-type structure with one water-water hydrogen bond replaced by a hydrogen bond to a helix carbonyl oxygen. Interestingly, in this case there is also evidence of these separations for the water-chain system, which may again be attributed to interactions with the ASP side chain groups, in this case with tightly coordinated water molecules. The second peak also shifts to a slightly shorter separation of 2.75 Å, which is more in line with the findings for the poly-glycine model.

5.3.3. Effect of changing the water density in the poly-glycine-aspartic acid model

As with the previous poly-GLY models the constraints imposed on the size of model system in order to allow for QM modelling methods to be employed mean that the model cannot support a water reservoir at either end. A predetermined number of water molecules must therefore be already present in the pore and thus the effect of varying their number must be checked.

Turning back to the rmsd plots shown in Figure 5.3 it is apparent that varying the density of the water molecules (from five to six to nine) has little effect on the average rmsd. Each value lies between 0.6 and 0.7 Å and all are within one standard deviation of each other. A possible link had previously been observed between the density of water in a poly-glycine channel and helix behaviour [14] , however in this case it would appear that the addition of positional restraints to backbone nitrogen atoms, and perhaps also the behaviour of the ASP side chains overrides any influence of the water density on the helical backbone behaviour.

The average atomic distributions of the water and the ASP oxygen-atoms have been used to investigate the ordering of water molecules along the poly-GLY-ASP channel *b*-axis, as shown in Figure 5.5. The peak distributions for each of the water molecules are shown on the bottom axis, whilst the six upper peaks represent the movement of the three ASP groups (with two peaks per ASP). As mentioned previously, the water-wire itself is only in existence for *ca.* 8% of the simulation, so the information in these plots is focused more on the interaction between the water molecules and the ASP side-chain residues. Concentrating first on the six water model [see Figure 5.5(b)], it is apparent that for the majority of the simulation, two of the water molecules [labelled $H_2O(1)$ and $H_2O(5)$, see Figure 5.1 for molecule labelling] are held in extremely tight configuration for the whole production run, as evidenced by the sharpness of the oxygen distribution peaks. $H_2O(1)$ participates in a strong hydrogen bond interaction with a backbone carbonyl oxygen and $H_2O(5)$ remains tightly coordinated to ASP3, again *via* hydrogen-bonding. $H_2O(6)$ diffuses *ca.* 10 Å (discussed further in Section 4.3.4) in the first 2.2 ps of the simulation, but then forms a strong hydrogen bond to $H_2O(1)$, resulting in a carbonyl/two water-molecule complex which exists for the remaining 25 ps of the MD. Similarly, once the PT process is complete,

$H_2O(4)$ forms an accepted hydrogen bond with ASP2, which again lasts for the remainder of the simulation. Molecules $H_2O(2)$ and $H_2O(3)$ do not form any sustained hydrogen bonds with the helix but remain 'trapped' between the ASP residues, thereby reducing their mobility as shown by their relatively narrow distributions in Figure 5.5(b).

Increasing the number of water molecules from six to nine alters the movement of the water molecules significantly. In this case $H_2O(2)$ and $H_2O(5)$ form lasting hydrogen bonds with ASP residues; these are the two narrow distributions in the plot [see Figure 5.5(c)]. All other water molecules are very mobile and migrate to the bottom third of the channel, indeed by the mid-point of the simulation seven of the nine water molecules reside in this portion of the channel. They form transient hydrogen-bonded clusters of varying numbers but none are long-lasting, apart from $H_2O(6)$ which hydrogen bonds to $H_2O(2)$. The migration results in 'dry' areas of the channel, punctuated only by $H_2O(5)$ and $H_2O(7)$, which, although do not coordinate to an ASP residue remain in the top section of the channel. This behaviour is not observed in the other two model systems. It is somewhat counterintuitive that the most water-dense model would result in the biggest dry areas, but this is perhaps due to the stability that larger water clusters can afford.

Reducing the number of water molecules in the channel from six to five has a noticeable effect on both its structure and dynamics [see Figure 5.5(a)]. In this case molecules $H_2O(2)$ and $H_2O(3)$ now participate in a constant hydrogen bonded chain with ASP1 which is formed within the first 2 ps, shown by relatively narrow peaks. Molecules $H_2O(1)$, $H_2O(4)$ and $H_2O(5)$ are all considerably more mobile, with $H_2O(1)$ travelling over 10 Å throughout the simulation. These three water molecules form transient hydrogen bonds with the ASP residues over the course of the simulation, but none are long-lived.

Turning to consider the movement of the ASP groups, there are there some similarities to be found in all the models. All three ASP residues interact with the water molecules in the central cavity. ASP3 reorientates itself during the geometry optimisation to hydrogen bond with a carbonyl group on a neighbouring helix in all of the models and remains in this position.

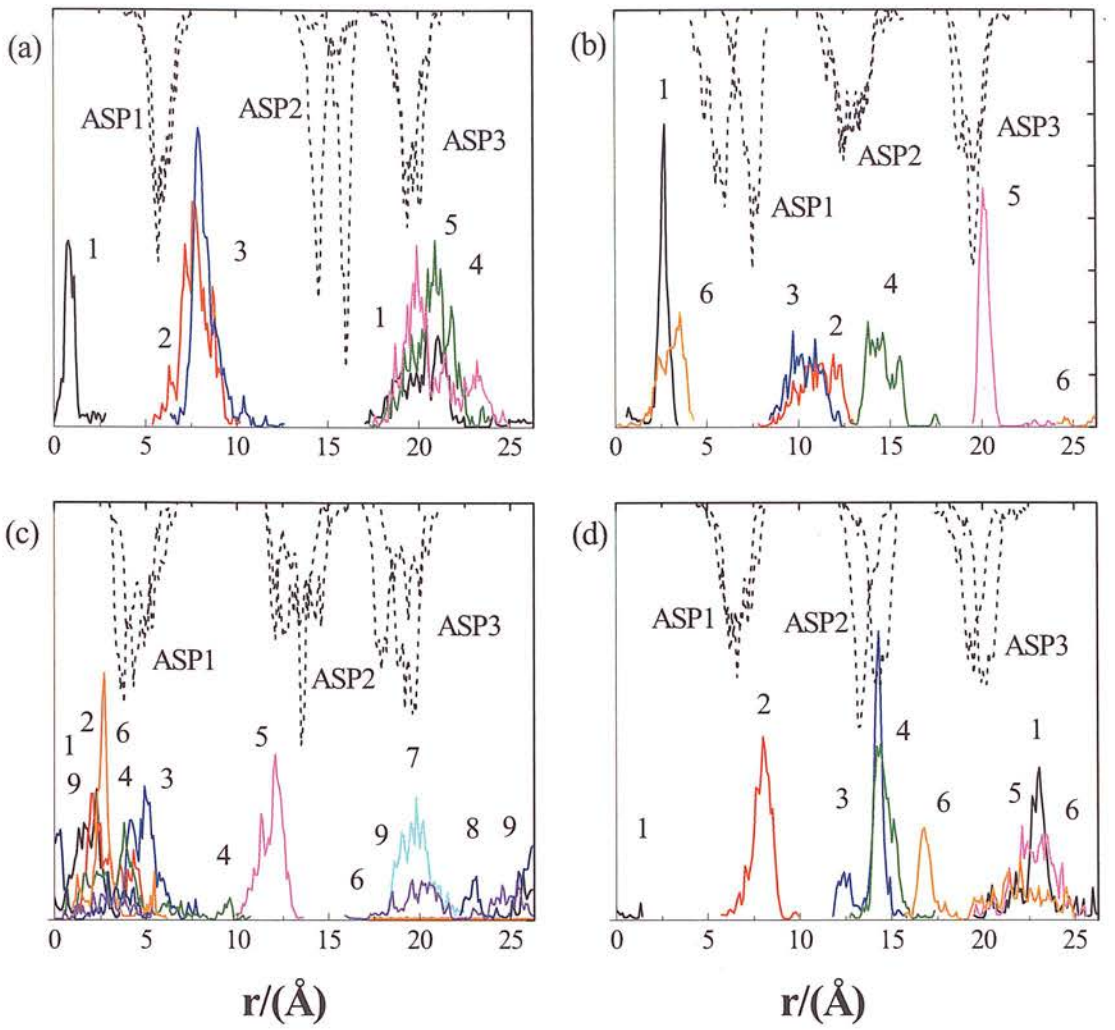


Figure 5.5 Average atomic distributions of water oxygen (coloured solid line) and aspartic oxygen (black dashed line) atoms modelled along the channel *b*-axis for the (a) five-, (b) six-, (c) nine-water-wire models. Also shown is the (d) six water chain model.

5.3.4. The PT pathway observed in the poly-glycine-aspartic model

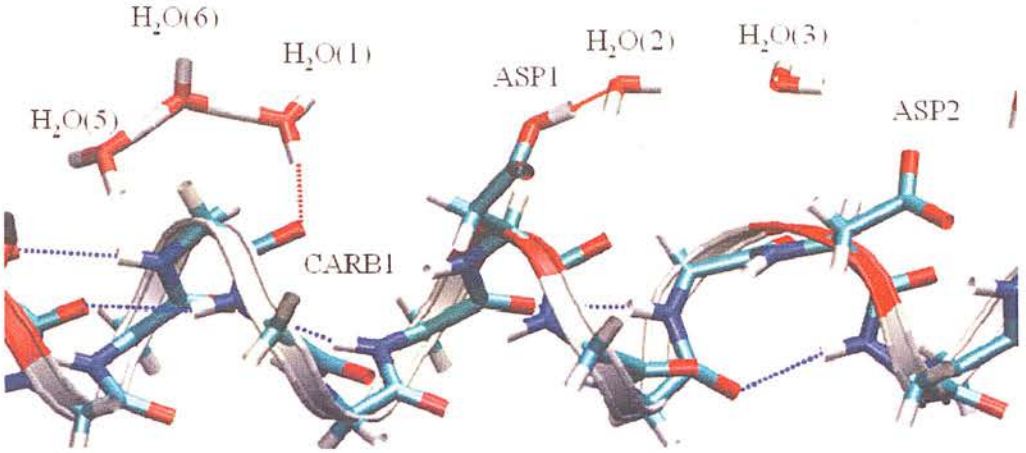
For all models a proton has been removed from ASP2 and placed on a water molecule in close proximity to the 'top' of the channel (*i.e.* above ASP3). For both the five and six water model $H_2O(5)$ becomes the oxonium, but the proton is added to $H_2O(8)$ in the nine water model. The six water model sees the proton travels over three

water molecules, a backbone carbonyl oxygen and one ASP residue to reprotonate ASP2 (*i.e.* the site from which it was removed). This equates to a distance of approximately 15.4 Å, which is over half the length of the channel, without any significant diffusion of the oxonium or [H_3O^+] ion. This is a greater distance than observed for the poly-GLY model, when the proton travelled a maximum of 10.5 Å. This is not surprising given the addition of the ASP residues to the helix scaffold: they provide both protonatable sites and anchor points for the hydrogen bonded water-wire. In addition these models contain a chemical gradient: the positive charge (H^+) recombines with the negative charge (COO^-). Proton transport itself occurs in three distinct steps:

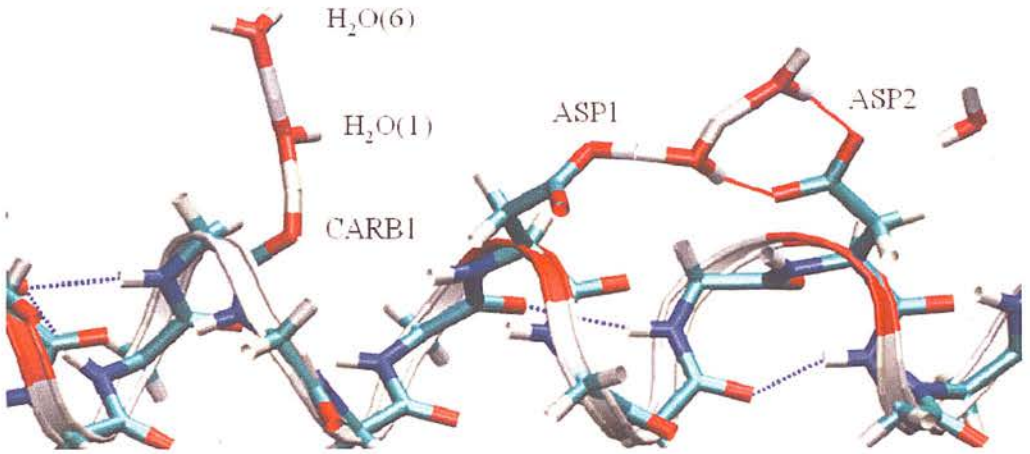
1. The excess proton is transferred *via* a series of hydrated Zundel [$H_5O_2^+$] cations in a 'standard' PT mechanism over two water molecules [$H_2O(5)$ and $H_2O(6)$] to reach $H_2O(1)$ [see Figure 5.6(a)].
2. The proton then moves to a backbone carbonyl oxygen [labelled CARB1, see Figure 5.6(b)] where it forms a stable complex with $H_2O(1)$ and $H_2O(6)$. This complex has a lifetime of 1.9 ps (3435 MD steps), which is equal to approximately 86% of the time taken for reprotonation of ASP2. Initially, however, this complex is not in the correct orientation to allow for transport to ASP1 and some reorganisation is necessary. In order to bridge the large gap between $H_2O(1)$ and ASP1 $H_2O(6)$ repositions itself *via* a 'slingshot' type movement. Strong hydrogen-bonding interactions exerted by the excess proton keep the complex intact throughout the period of time needed for $H_2O(6)$ to traverse almost 8 Å [detailed in Figures 5.6(a)-(c)]. This is the rate-determining step or 'bottleneck' of the reaction.

Once the correct positioning of $H_2O(6)$ is achieved [see Figure 5.6(d)], step three sees the proton released from the CARB1 and transferred *via* $H_2O(1)$ and $H_2O(6)$ to ASP1, which loses a proton to $H_2O(2)$ which in turn finally donates to ASP2 in a cascade of almost instantaneous PT steps. At this point a hydrogen bonded wire is formed which sees ten oxygen atoms acting cooperatively [see Figure 5.6(d)]. Once reprotonation occurs this extended complex quickly breaks down and is not observed again; the water molecules are then able to diffuse throughout the channel fairly freely [see Figure 5.6 (e)-(f)].

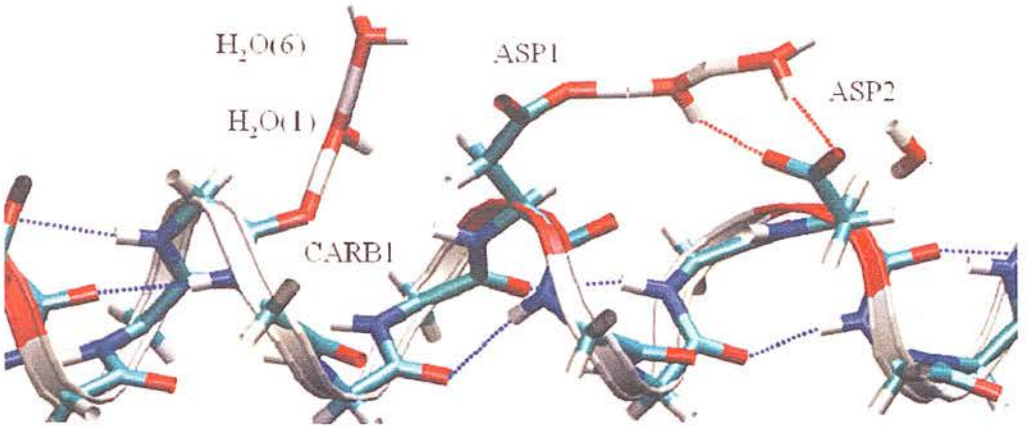
(a)


H2O(6)
H2O(1)
H2O(5)
ASP1
H2O(2)
H2O(3)
ASP2
CARB1

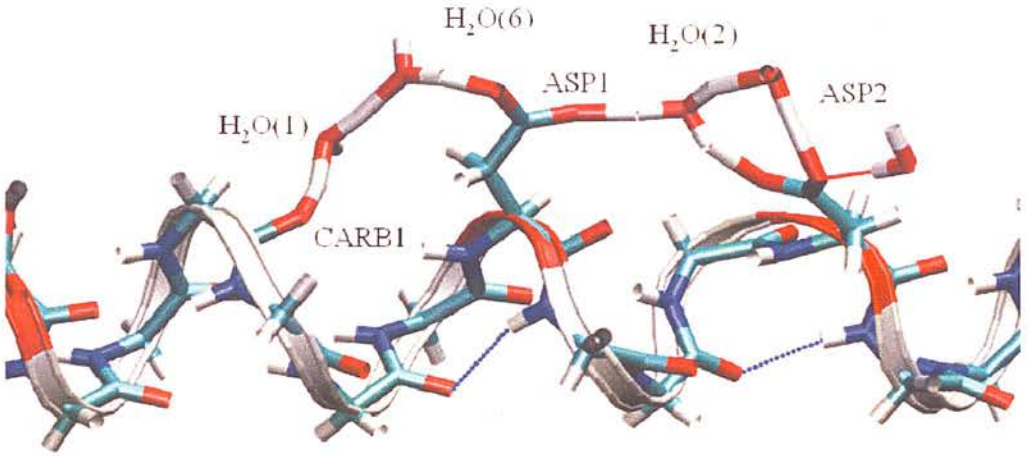
(b)


H2O(6)
H2O(1)
CARB1
ASP1
ASP2

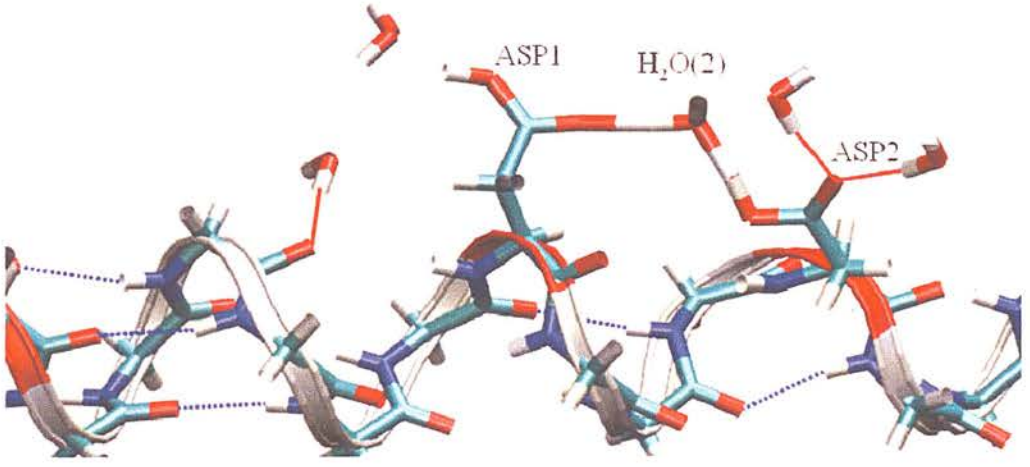
(c)


H2O(6)
H2O(1)
ASP1
ASP2
CARB1

(d)


H2O(6)
H2O(2)
ASP1
ASP2
H2O(1)
CARB1

(e)


ASP1
H2O(2)
ASP2

(f)

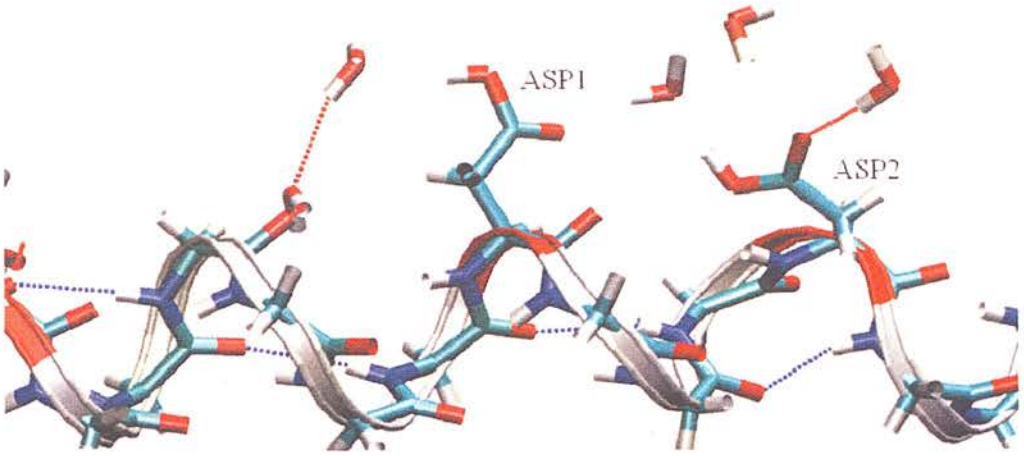


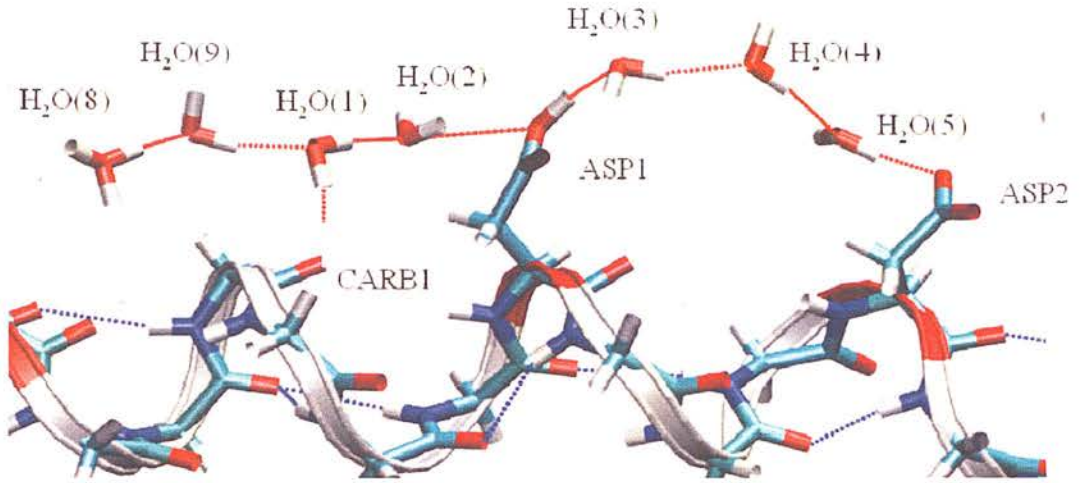
Figure 5.6 Schematic of the proton transport process for the six water-wire model. Step 1 is detailed in (a), step 2 in (b), (c) and (d) and finally step 3 is shown in (e) and (f).

The addition of three water molecules to the channel does not change the overall PT pathway, but does lead to a few important changes in the process. Initially there is pre-organisation of the water molecules to form a vast hydrogen-bonded system leading from $H_2O(8)$ through to ASP2, comprising seven water molecules, one carbonyl oxygen and two ASP residues [see Figure 5.7(a)]. This complex lasts for 65% of the time taken for PT to occur (*i.e.* 0.7 ps) but breaks down *via* the loss of two water molecules [$H_2O(8)$ and $H_2O(9)$] to form a smaller complex (which will be discussed below). The pathway can be summarised as follows:

1. Step one sees 'standard' PT from $H_2O(8)$ to $H_2O(1)$ *via* $H_2O(9)$, where it again remains for a relatively long period of time [0.64 ps, see Figure 5.7(a)]. The proton is not transferred to the CARB1 in this case, but rather remains on $H_2O(1)$ which in turn is tightly coordinated to CARB1. The addition of a water molecule to this section of the wire means there is no longer a gap to bridge to ASP1 and as such there is now a fully organised water-wire which leads from CARB1 to ASP2, *via* five water molecules and ASP1 [see Figure 5.7(b)]. This complex lasts for a further 22% (0.2 ps) of the time taken to reprotonate ASP2.
2. In this point ASP1 cannot accept the excess proton without first losing its existing one, [see Figure 5.7(b)]. Thus step two, or the rate determining step

(0.77 ps of 1.1 ps) in this reaction is the loss of the ASP1 proton. After this occurs the reaction again proceeds almost instantaneously from ASP1 to ASP2 *via* $H_2O(3)$, $H_2O(4)$ and $H_2O(5)$. Note that the whole process takes almost half the time needed for that of the six water model.

(a)



(b)

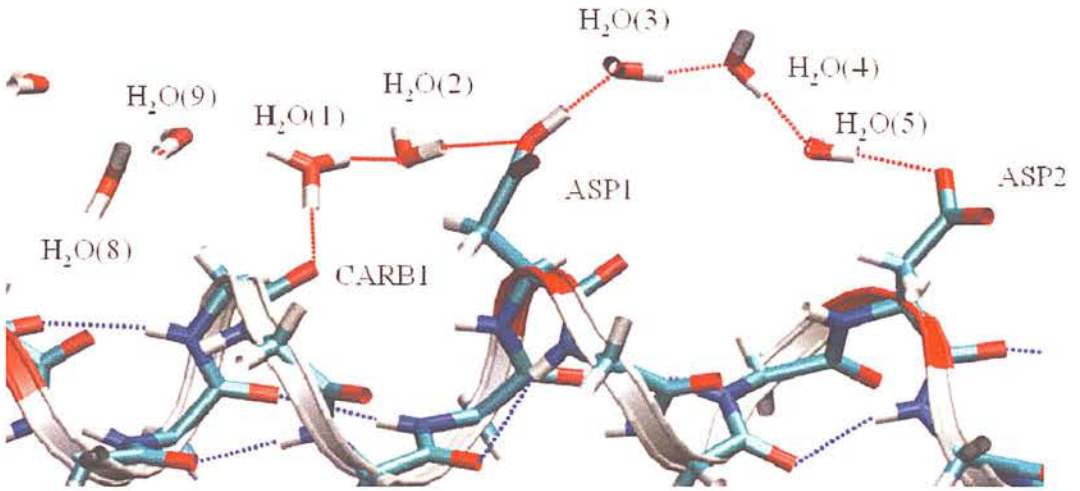
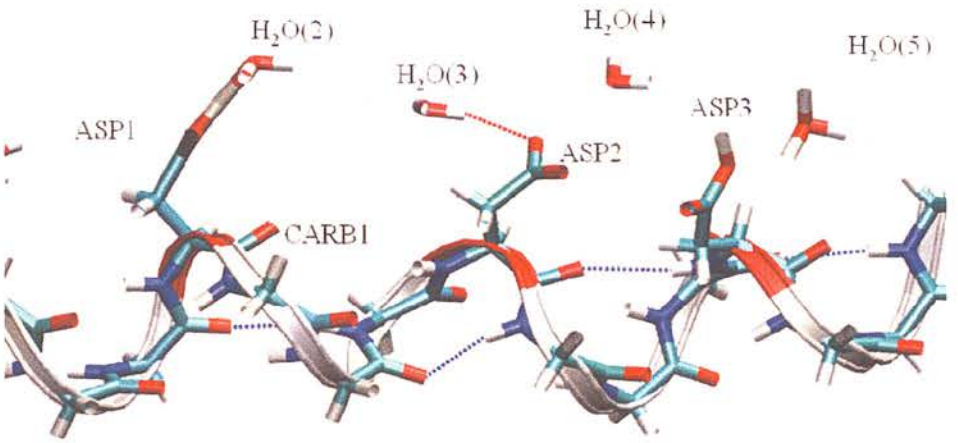


Figure 5.7. The proton transport pathway in the nine water model. (a) Step 1 and (b) Step 2

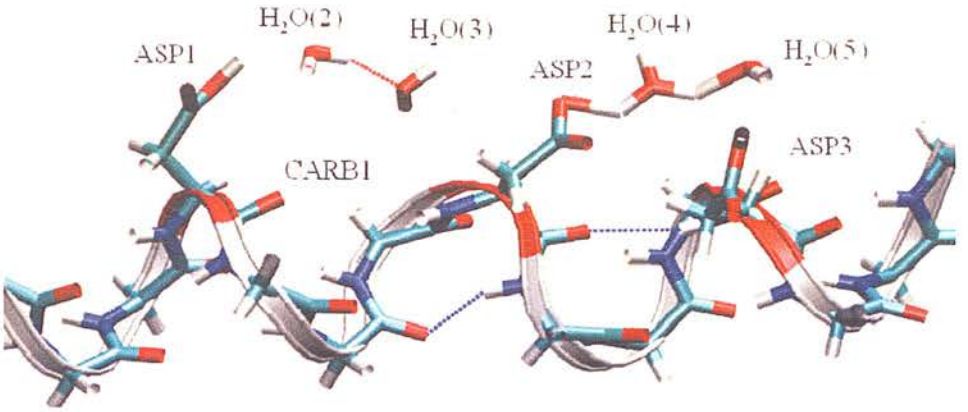
Reducing the number of water molecules in the channel to five greatly changes the PT pathway. The water molecule removed was $H_2O(6)$, which has played an

integral role in the transport processes of the two models already discussed. The proton cannot reach the helix carbonyl backbone or ASP1 and is forced to travel in the opposite direction over one further water molecule directly to ASP2 [see Figure 5.8(a)-(c)]. This method necessitates some diffusion of the initial oxonium ion (*ca.* 2 Å), simply because there are no water molecules or ASP residues in close proximity. The oxonium ion thus diffuses past ASP3, forming transient hydrogen bonds only, before transferring *via* a standard PT mechanism to $H_2O(4)$ and finally to ASP2, in a time of 0.2 ps. This does raise the question of why, for the six and nine water models, the proton chooses to take the longer route back to ASP2. The proton travels only 7 Å in the five water model, in comparison with almost 16 Å for the six and nine water models, and takes a fraction of the time at 0.2 ps compared to 2 ps and 1.2 ps for the six and nine water models, respectively. This points to multiple possible pathways for transport.

(a)



(b)



(c)

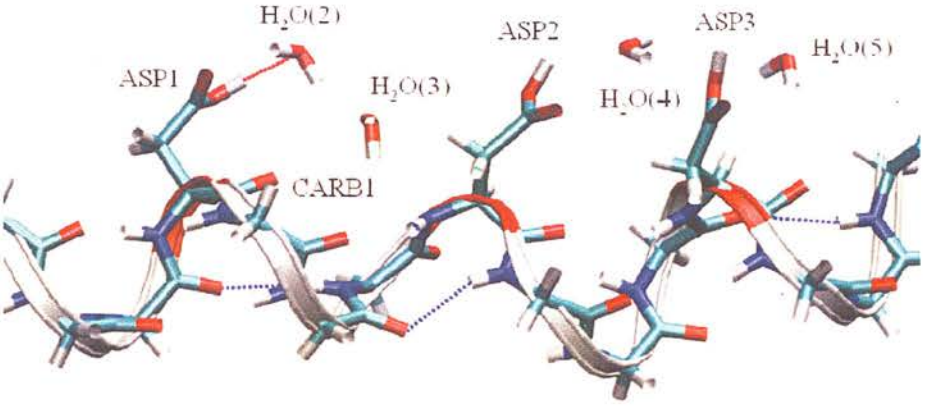


Figure 5.8 (a) shows the first step in the PT process for the five water-wire model. The oxonium ion [$H_2O(5)$] diffuses to reach $H_2O(4)$ as shown in (b) which then reprotonates ASP2 and the briefly formed water-wire breaks down in (c).

The RDFs of the oxygen atom of the initial oxonium ion go some way to provide a rationalisation for this behaviour. Figure 5.9(b) shows the $g_{OO}(r)$ plot of the initial oxonium oxygen to the two nearest neighbours for the six water model. The solid line indicates the nearest neighbour for the direction in which PT occurred, *via* $H_2O(6)$ and therefore the longer of the two possible routes. The dashed line shows ASP3 which is directly below the water molecule, *i.e.* the gateway to the shorter path. $H_2O(6)$ is the closer of the two options and as such is the favoured next step for the proton. The

corresponding RDF for the five water model [Figure 5.9(a)] is again easy to interpret and shows that there is no clear path for the proton to follow as the closest possible oxygen is over 3.5 Å away. The proton 'chooses' to follow the shorter, more direct path and diffuses towards ASP3. The RDF for the nine water model, however, provides another insight into the PT pathway. The initial $g_{OO}(r)$ plot shows that the ASP3 and $H_2O(9)$ are equidistant yet the proton still chooses to travel *via* the longer path back to the ASP2. This shows the extent of influence of the pre-organisation of the water-wire. The seven water-molecule, one carbonyl and two ASP complex formed at the very beginning of the simulation provides a much easier path for the proton to travel along. Diffusion of the oxonium ion, which provides a shorter, more direct path to reprotonation, is not preferred to traditional proton transfer along a series of hydrogen-bonded water molecules.

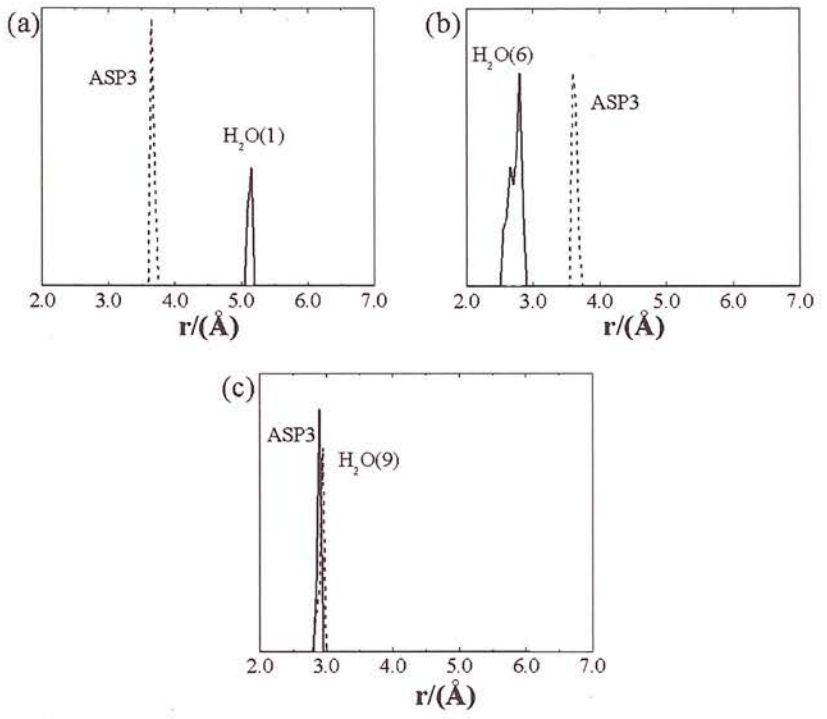


Figure 5.9. Time-dependent RDFs, $g_{OO}(r)$, of the initial oxonium oxygen to its two nearest neighbours, one water molecule (solid black line) and one aspartic acid (dashed black line) for the (a) five- , (b) six- and (c) nine-water-wire models.

5.3.5. Configuration of the aspartic acid head groups

An ASP carboxyl group can exist in more than one orientation, the lowest energy configuration being that where the OCOH torsional angle is 0° (or 360°), which occurs when the -C=O and –OH groups are in line (see Figure 5.10, obtained by performing a 6-31G*/b3lyp potential energy surface scan on CH_3COOH using the Gaussian03 software package [16]). A second orientation of the hydrogen is possible, where the OCOH torsional angle is 180°, but it is relatively rare, as this local minimum is some 30 $kJmol^{-1}$ higher in energy than the global minimum described above. The energy barrier between the two configurations is *ca.* 35 $kJmol^{-1}$, meaning that flipping between the two cannot be achieved with ease.

Figures 5.11(a) and (b) follow the OCOH angle of the transferring proton and aspartic group one for the six water model, with (a) tracking the entire simulation and (b) showing the first 4 ps only. The first *ca.* 2 ps show the angle of approach of the water hydrogen which will transfer to the aspartic acid group (in effect the H...OCO angle). The rest of the plot shows the resulting HOCO torsional angle after transfer has occurred. The H...OCO angle is predetermined as a result of hydrogen bonding of the ASP group to $H_2O(2)$ [see Figure 5.6(e)], which is part of a wider hydrogen bonded chain to ASP2. ASP1 is protonated from below at an angle of approximately 110°, which lies on the right hand side of the potential energy curve shown in Figure 5.10. Once it falls into this energy well it cannot overcome the barrier and return to its preferred orientation and retains the reverse configuration, shown by an angle of 180° in Figure 5.11 (a) and (b) and a schematic in Figure 5.12 (a).

Figure 5.11(d) shows a similar plot for the nine water model. In this case the same aspartic residue (ASP1) is protonated during the transfer process from below, at an angle of 240°. In this case, however, the proton is able to overcome the barrier and revert to its minimum energy arrangement, *i.e.* an OCOH torsional angle of 0°. There is increased rotation of the ASP1 side chain in this transfer process and this, coupled with the hydrogen-bonding influences to water molecules, facilitates the ASP1 OCOH group returning to its preferred configuration.

The path for PT in the five water molecule model does not directly involve ASP1, but ASP2 exhibits some interesting behaviour, as shown by its torsional angle plotted in Figure 5.11(c). In this case, ASP2 is reprotonated directly from above at an angle of 0° but after some time flips to the reverse configuration and forms a sustained hydrogen-bond with a backbone carbonyl oxygen [see Figure 5.12(d)]. The PT process occurs at a greatly increased rate, resulting in an increase in rotational energy of the ASP2 group after collision with the transferring proton. This rotation allows the ASP group to come into close contact with the neighbouring carbonyl oxygen and with this an opportunity to form a hydrogen-bond. This additional interaction must provide an energetically favourable arrangement which is difficult to break.

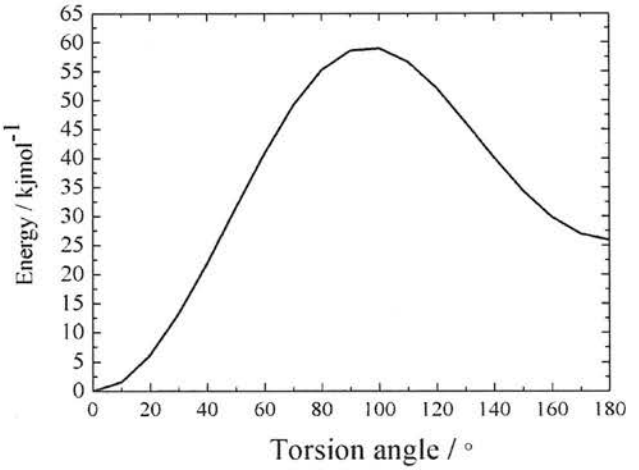


Figure 5.10 Potential energy surface of CH_3COOH from an angle of 0 to 180°.

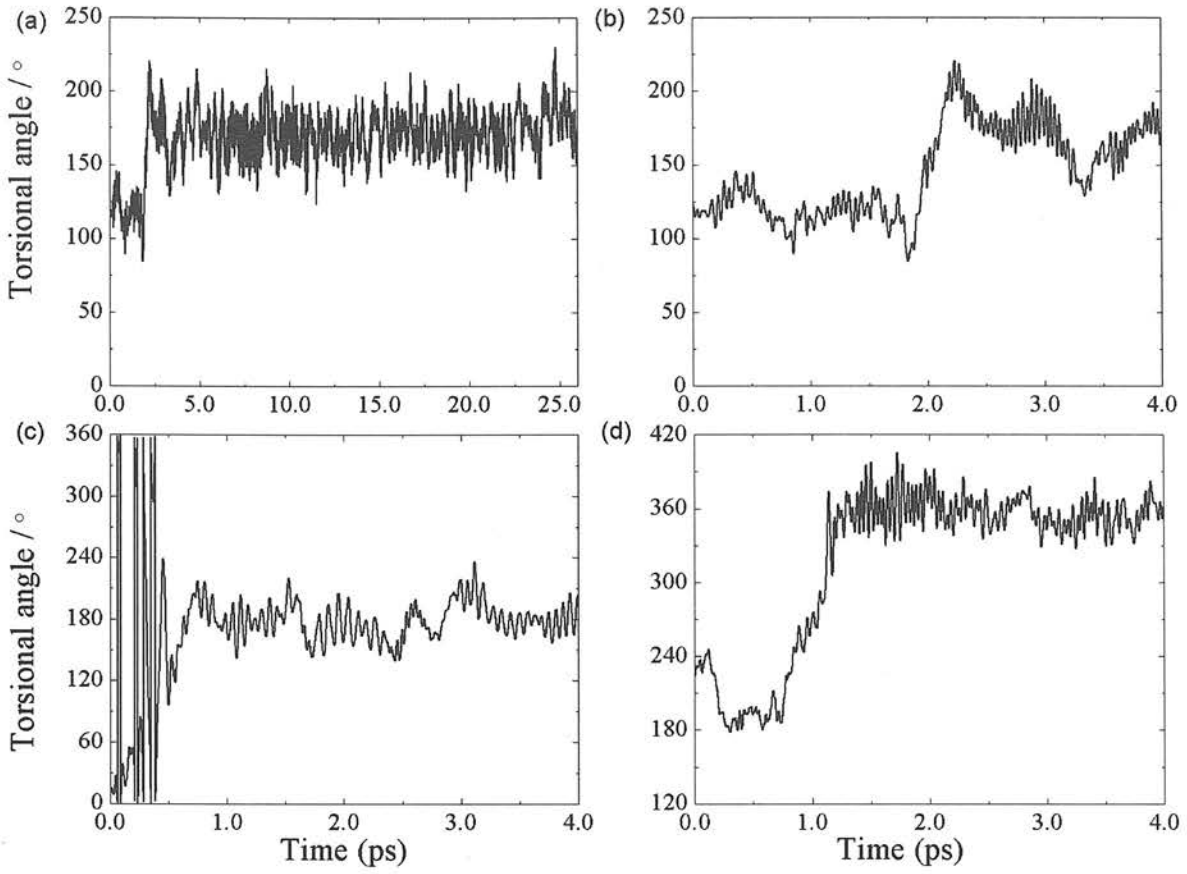
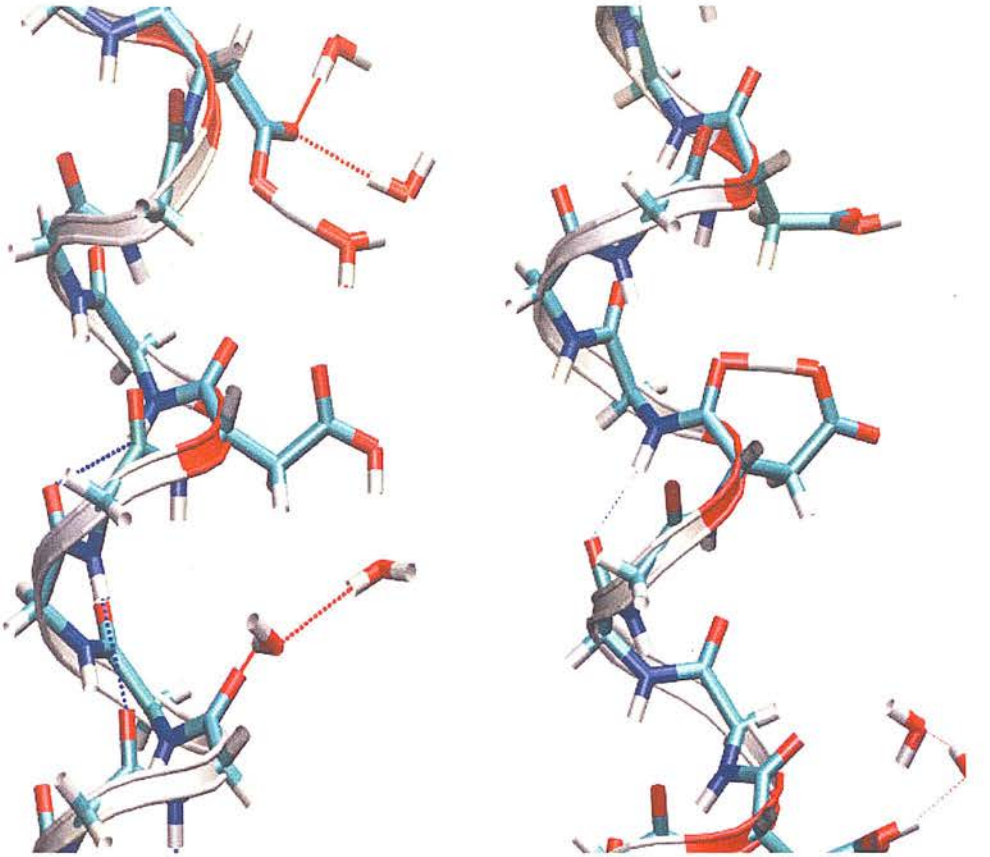


Figure 5.11 Plots of the OCOH torsional angle for the six water-wire model (ASP1) for (a) the entire MD simulation and (b) a snapshot of the first four picoseconds. Also shown are snapshots for the (c) five- (ASP2)and (d) nine-water-wire models (ASP1).



(a) (b)

Figure 5.12 Representation of the reverse configuration adopted after reprotonation of (a) ASP1 of the six water-wire model and (b) ASP3 of the five water-wire model.

5.4 Conclusions

The simplified poly-GLY model system used to investigate PT through alpha-helical channels has been modified to include three aspartic acid residues, which are capable of direct interaction with the water wire. In addition, a chemical driving force has been created by the deprotonation of an ASP group and subsequent placing of the excess hydrogen on a water molecule to create an oxonium. The introduction of this chemical driver greatly increases both the rate of PT and also the distance travelled by the proton in comparison with the three previous models; the ASP group is reprotonated in less than 2 seconds over a distance of 15 Å.

The removal of a proton and placement on a water molecule within the water chain does not, in this case, exert a significant influence over the behaviour of the water molecules There is some evidence of increased density at shorter O-O separations compared with the baseline 6 water model, but there is no obvious structuring of the water solvation shell. This would suggest that the influence of the polar aspartic groups is much greater than that of the excess proton.

The ASP groups are shown to participate directly in the proton transfer process. They serve as a 'bridge' between chains of water molecules as protons are exchanged *via* the labile carboxylic acid head groups; C=O accepts a proton, whilst the hydrogen of the CO-H group is lost. The ASP groups also provide anchor points for the water molecules, with hydrogen bonds lasting the entire length of the simulation.

A reaction path for PT for each of the three models (*i.e.* five, six and nine waters) has been observed. The proton moves in the same direction for both the six and nine water molecules. The rate limiting steps, however, are slightly different. The nine water model provides a continuous, unbroken chain of water molecules from the initial positioning [on $H_2O(6)$] of the proton to ASP2. This route is significantly longer than simple diffusion of an oxonium ion to reprotonate the ASP residue (as in the five water model discussed below). The rate limiting step in this process is the loss of the existing ASP1 proton. The removal of three water molecules creates a break in the chain and in this case the 'bottleneck' for the reaction is the time taken for diffusion of a water molecule to form a bridge to the ASP group.

The pathway for proton transport in the five water model is significantly different. RDFs at the beginning of the simulations show that there is no pre-organised water-wire for this model, so the initial oxonium diffuses some distance in order to facilitate reprotonation of the ASP group. The density of water molecules has thus been shown to affect the direction of PT. When there is sufficient density to form a pre-organised hydrogen bonded network, the proton will choose to travel *via* this network, even if it is the longer route.

5.5. Bibliography

1. Thompson, M.A., *ArgusLab 4.0.1*, Planaria Software LLC: Seattle, WA.

2. Accelrys, *Materials Studio*. 2002: San Diego.

3. The CP2K Developers' Group. 2008, http://cp2k.berlios.de

4. Becke, A.D., *Density-functional exchange-energy approximation with correct asymptotic behavior.* Physical Review A, 1988. **38**(6): p. 3098-3100

5. Lee, C., W. Yang, and R.G. Parr, *Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density.* Physical Review B, 1988. **37**: p. 785-789

6. Goedecker, S., M. Teter, and J. Hutter, *Separable dual-space Gaussian pseudopotentials.* Physical Review B (Condensed Matter), 1996. **54**(3): p. 1703-1710.

7. Broyden, C.G., *The Convergence of a Class of Double-rank Minimization Algorithms.* Journal of the Institute for Mathematics and Applications, 1970. **6**: p. 222-231.

8. Fletcher, R., *A new approach to variable metric algorithms.* Computer Journal,, 1970. **13**: p. 317-322.

9. Goldfarb, D., *A family of variable metric methods derived by variational means.* Mathematics of Computation, 1970. **24**: p. 23-36.

10. Shanno, D.F., Mathematics of Computation, Vol. 24, pp 647-656 1970, *Conditioning of quasi-Newton methods for function minimization.* Mathematics of Computation, 1970. **24**: p. 647-656.

11. Grimme, S., *Semiempirical GGA-type density functional constructed with a long-range dispersion correction.* Journal of Computational Chemistry, 2006. **27**(15): p. 1787-1799.

12. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics.* Journal of Molecular Graphics, 1996. **14**: p. 33-38.

13. Brewer, M.L., U.W. Schmitt, and G.A. Voth, *The Formation and Dynamics of Proton Wires in Channel Environments.* Biophysical Journal, 2001. **80**: p. 1691-1702.

14. Shepherd, L.M.S. and C.A. Morrison, *Simulating Proton Transport through a Simplified Model for Trans-Membrane Proteins.* Journal of Physical Chemistry B, 2010. **114**: p. 7047-7055.

15. Smart, O.S., et al., *HOLE: A program for the analysis of the pore dimensions of ion channel structural models.* Journal of Molecular Graphics, 1996. **14**: p. 354-360.

16. Frisch, M.J., et al., *Gaussian 03*. 2004, Gaussian Inc: Wallingford, CT.

Chapter 6:

Conclusions

Final conclusions and potential options to extend the research

6.1. Conclusions

The aims of establishing whether or not a greatly simplified model for a trans-membrane protein channel can be designed and built, and of whether or not this simplified model could support fast proton transport have been successfully met.

A basic poly-glycine scaffold has been constructed, shown to be sufficiently stable to withstand the molecular modelling process and also to be capable of supporting Grotthuss-style fast proton transport. The presence of an excess proton induces long-range co-operativity along the water-wire and the formation of a distinct second solvation shell. Analysis of the mechanism for fast proton transport in this most basic of the model systems used has highlighted some important deviations from the accepted bulk water mechanism. The number of hydrogen atoms directly co-ordinated to the O* in bulk water is four, reducing to three before successful proton transport occurs. The narrow channel environment allows only for a co-ordination sphere of three around O* so the loss of the fourth hydrogen, which is critical for successful transfer in bulk water, is not a necessary pre-cursor to transfer in the channel environment. The formation of a Zundel complex, however, is required prior to transfer in both the bulk and channel environments.

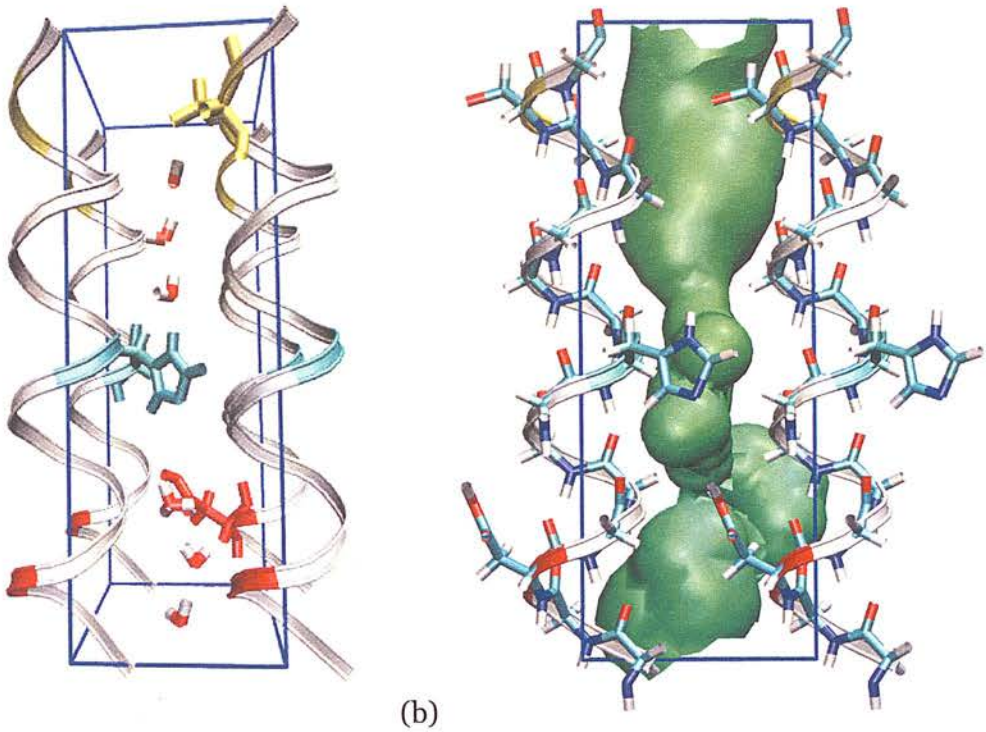
The addition of polar side chain residues to the poly-glycine scaffold results in breaks in the water-wire greatly limiting the length of wire formed and thus the extent of proton transport. Extended delocalisation of the excess proton is possible *via* co-ordination to polar serine and aspartic side chains and also to helix carbonyl oxygen atoms. This is perhaps more true to real life where water molecules are thought to form bridges between important residues in trans-membrane protein channels such as the D-pathway of Cytochrome C Oxidase. In the model systems where an excess proton was included in the water-wire to form a charged system there was, however, no direct participation of these residues in proton transport. The mechanism for PT was the same as observed in the basic poly-glycine model. The poly-glycine-aspartic model has been used to investigate the 'special-pair dance' seen before PT in bulk water. Increased 'rattling' events in the narrower –glycine and –serine channels may mask this

phenomenon, but the width of the –aspartic channel allows for the formation of an Eigen-style cation which results in the proton remaining stationary for prolonged periods. In these periods flipping of the identity of the closest oxygen, *i.e.* O_{next} has been observed, indicating that as soon as there is more than one possible plane of movement for the proton this becomes an integral feature of the transfer process.

A different approach was used in the final model system, whereby a proton was removed from a serine group and placed on a water molecule. This creates a chemical gradient and thus 'driver' for the process. The preferred pathway of the proton was determined to depend on the extent of pre-organisation of a hydrogen-bonded network. If there is an existing network present, the proton will choose to travel along this pathway *via* Grotthuss 'hopping' even if this provides a much longer route than simple diffusion of water molecules. In these model systems the aspartic residues have been shown to participate in PT through the loss of the –C-OH proton and gain of a proton to the C=O group, resulting in a reversal of the carboxylic acid head group.

The flexibility of the model system created means that the side chain groups can be easily changed in order to model specific systems of interest. One such system is the 'catalytic triad' – a set of three amino acid residues which work together to hydrolyse peptide bonds. The three residues of interest are serine, aspartic acid and histidine. Although it is not possible to directly replicate the process of hydrolysis within our models, these head groups have labile protons and are thus active for proton transport. As such they form an interesting model system for study.

A system containing the three active amino acid residues has been created by the addition of an aspartic acid, a histidine and a serine residue to the poly-glycine scaffold. This is then coupled to a six-membered water-wire. A model of the simulation cell with the amino acids coloured by type is shown below in Figure 6.1(a) alongside a representation of the channel constructed using the HOLE program.



(a) (b)

Figure 6.1 (a) Model of the catalytic triad simulation cell and (b) HOLE representation of the channel

Chapter 7:

Appendices

7.1 Papers published and in preparation

- Shepherd, L.M.S. and C.A. Morrison, *Simulating Proton Transport through a Simplified Model for Trans-Membrane Proteins.* The Journal of Physical Chemistry B, 2010. **114**(20): p. 7047-7055.

- Competing pathways for proton transport through a poly-glycine-aspartic channel: an *ab initio* molecular dynamics study (in preparation)

7.2 Conferences and Summer Schools attended

7.2.1. Conferences

- 12th Annual Density Functional Theory Conference (Amsterdam 2007)
- Faraday Discussion 141: Water: From interfaces to the bulk (Edinburgh 2008)
- SCOTChem Computation Chemistry Symposium (Edinburgh 2009)
    - 2nd Place Poster Prize and Flash Presentation
- 42nd IUPAC World Chemistry Conference (Glasgow 2009)
    - Poster and Flash Presentation

7.2.2. Summer Schools

- Rare Event Modelling Course (Oxford 2008)
- CCP5 Summer School (Sheffield 2008)
- UK Theoretical Chemistry Summer School (Oxford 2008)